AFHRL-TP-83-16

PROCEEDINGS

24th ANNUAL CONFERENCE

MILITARY TESTING ASSOCIATION

COORDINATED JOINTLY BY
AIR FORCE HUMAN RESOURCES LABORATORY
AND
USAF OCCUPATIONAL MEASUREMENT CENTER

★ ★ ★ ★ ★

SAN ANTONIO, TEXAS
1-5 NOVEMBER 1982

COMPONENT PART NOTICE

THIS PAPER IS A COMPONENT PART OF THE FOLLOWING COMPILATION REPORT:

(TITLE): Proceedings: Annual Conference of the Military Testing Association (24th) Held at San Antonio, Texas on 1-5 November 1982.

(SOURCE): Air Force Human Resources Lab. Brooks AFB, TX

TO ORDER THE COMPLETE COMPILATION REPORT USE AD-A126 554.

THE COMPONENT PART IS PROVIDED HERE TO ALLOW USERS ACCESS TO INDIVIDUALLY AUTHORED SECTIONS OF PROCEEDINGS, ANNALS, SYMPOSIA, ETC. HOWEVER, THE COMPONENT SHOULD BE CONSIDERED WITHIN THE CONTEXT OF THE OVERALL COMPILATION REPORT AND NOT AS A STAND-ALONE TECHNICAL REPORT.

THE FOLLOWING COMPONENT PART NUMBERS COMPRISE THE COMPILATION REPORT:

COMPONENT PART NOTICE (CON'T)

AD#:		TITLE:
	P000 828	Definition, Classification, and Measurement of Turnover in Civilian and Military Contexts.
	P000 829	Process Models of Turnover Within an Open-Systems Context.
	P000 830	Officer Surveys - Canadian Forces.
	P000 831	Officer Occupational Surveys in the Marine Corps.
	P000 832	The Use of Job Analysis Information in Assigning Managers to Positions in a Diagnostic Organizational Simulation.
	P000 833	The Use of Job Analysis Inventories for Developing Improved Selection Systems.
	P000 834	Job Analysis and Task Oriented Ratings.
	P000 835	The Use of Job Analysis Inventories in Job Evaluation.
	P000 836	The Use of Task Based Job Analysis Data for Developing Performance Evaluation Systems.
	P000 837	Enlistment Motivation in the All-Volunteer Force Environment: A Review of Major Surveys.
	P000 838	The 1982 DA (Department of the Army) Survey of Personnel Entering the Army.
	P000 839	Systems Approach to Training - A Management Model.
	P000 840	Analysis and Evaluation of Command Group Operations.
	P000 841	Implementing the Criterion Referenced USAF (United States Air Force) Apprentice Knowledge Test Program.
	P000 842	Performance Appraisers as Test Developers.
	P000 843	Validation of the IRJ (Instantaneous Report of Judgements) Procedure for Analyzing the Rating Process.
	P000 844	Problem Solving and Student Motivation for Classroom Instructors.
	P000 845	Armored Fighting Vehicle Identification Training: A New Perspective.
	P000 846	Officer as a Counselor in Eliminating Self-Defeating Behavior.
	P000 847	Combining Results of Independent Research in Tank Crewman Performance.
	P000 848	Testing a Half Million Trainees: AFMET (Air Force Medical Evaluation Test) Project.
	P000 849	Maintenance Performance.
	P000 850	Validation Studies of the Belgian Armed Forces Reserve Officer Selection System.
	P000 851	Using Occupational Surveys to Develop Air Force Specialty Training Standards.
	P000 852	Subjective Appraisal as a Feedback Tool.
	P000 853	Implementation of an Air Force Occupational Research Data Bank.
	P000 854	Predicting Trainability of M1 (Abrams) Tank Crewmen.
	P000 855	United States Army Advanced Medic (91B30) Training: An Iterative Decision Method Application.
	P000 856	USAMPS (United States Army Military Police School) Test Validation Procedures.

AD:	P000 857	TITLE:	Development of the Army Research Institute Interest Survey (ARIIS).
	P000 858		Predictive Validity of Surrogate Measures of Job Performance for Navy GENDETs (General Detail Personnel).
	P000 859		The CODAP80 'RANDOM' Procedure.
	P000 860		Predicting Attrition in the Army Initial Entry Rotary Wing Course.
	P000 861		The Graduates' View of Naval ROTC (Reserve Officers Training Corps).
	P000 862		Organizational Maintenance Performance System.
	P000 863		A Comparison of Two Methods for Assessing Task Criticality.
	P000 864		The Jackknife: Its Application to Test Equating.
	P000 865		Disciplinary Penalties as Seen by Soldiers of the German Federal Armed Forces.
	P000 866		Military Eligibility and Participation in the All-Volunteer Force.
	P000 867		Instructor Attitudes Toward Instructional Systems Development and Performance Based Testing.
	P000 868		An Iterative Decision Method for Selecting Medical Tasks for Training.
	P000 869		Selection of Intelligence Analysts.
	P000 870		Development of Tank Platoon Gunnery Tactical Standards Through Simulation.
	P000 871		Predicting Subsequent Service Potential for Former Army Prisoners.
	P000 872		Personality Traits of U.S. Army Prisoners.
	P000 873		Career Attitudes of ROTC (Reserve Officers Training Corps) Cadets and Collete Students.
	P000 874		The Matrix Format for Task Inventories: A First Look.
	P000 875		Stabilized Gunnery Training Techniques.
	P000 876		Assessment of the Hometown Recruiter Assistance Program.
	P000 877		Computer Assisted Training for Letter Sorting Machine Operators.
	P000 878		Improved Course Test Design and Development.
	P000 879		Mission Based and ISD (Instructional Systems Development) Based Components of Task Criticality Judgements.
	P000 880		Does Measurement Hurt or Help a Sacrosanct System?
	P000 881		Identification of Rating Policies in Training Emphasis Task Factor Data.
	P000 882		A Comparison of Two Basic Russian Teaching Methodologies.
	P000 883		Providing the Man in the Back Seat.
	P000 884		Comparison of the Coast Guard's Selection & Classification Batteries.
	P000 885		Using Generalizability Theory to Assess Interrater Reliability of Contract Proposal Evaluations.
	P000 886		Using Automated Personality Test Interpretation for Security Screening.
	P000 887		Importance of Personal Characteristics to Job Performance.
	P000 888		Enlisted Military Selection: Impacts of Changing Aptitude Standards Since 1940.
	P000 889		Differences on Personality Measures Related to Recruit Attrition.

AD:	P000 890	TITLE:	Human Performance Requirements in C3I (Command, Control, Communications and Intelligence) Systems and Their Implications in System Design.
	P000 891		Organizational and Non-Organizational Influences on Attitudes Toward Work.
	P000 892		Assessing Motivational Skill Deficiencies in Military Trainees.
	P000 893		Issues Confronting the Design of a New Battle Drill Training System.
	P000 894		A Method for Improving Soldier's Manuals.
	P000 895		Retention of Armor Procedures: A Structural Analysis.
	P000 896		The Relationship Between Farming Activities, Physical Fitness and Body Fat.
	P000 897		A Comparison of Strategies for Improving Stress-Coping Responses.
	P000 898		Development of a New Form for Teacher Appraisals of Leadership.
	P000 899		Use of Delphi-Based Interdisciplinary Goals for Evaluating Course Outcomes.
	P000 900		Job Descriptions Developed from Groups X Module (GXM) Analyses.
	P000 901		Using Rating Scales to Determine Aptitude Requirements of Army Systems.
	P000 902		Research and Development for the JOIN (Joint Optical Information Network) System.
	P000 903		Geometric Radar Symbology: Static Cathode Ray Tube Testing.
	P000 904		Where Angels Fear to Tread — The Impact of CBI (Computer Based Instruction) Upon Instructors.
	P000 905		Women in the Navy's Civilian Skilled Blue-Collar Workforce.
	P000 906		Transition Socialization Processes in the U. S. Marines.
	P000 907		A Redesigned Performance Appraisal System for Noncommissioned Ranks in the Canadian Armed Forces.
	P000 908		Revision of the Organizational Assessment Package System: Improvements in Assessing Air Force Organizations.
	P000 909		Retrained Airmen: Volunteers versus Non-Volunteers.
	P000 910		Evaluating Individualized and Group Instruction Programs for Technical Associative Structure.
	P000 911		Management Information System for Processing USCG (United States Coast Guard) Class 'A' School Requests.
	P000 912		Job Difficulty Data as an Indicator of Job Complexity.
	P000 913		Benchmarking Occupational Survey Task Factor Data.
	P000 914		A Task Level Inventory for Describing Job Reading.
	P000 915		How To Display Data Badly.
	P000 916		A Personnel Test of Firefinder Radar Training, Trainer, and System.
	P000 917		The NROTC (Naval Reserve Officers Training Corps) Instructors Seminar: Preparing Officers for College Teaching.
	P000 918		Examinee and Accession Quality: Past and Present.
	P000 919		Introduction of Trade and Lifestyle Videotapes (TLVs) into a Canadian Forces Vocational Counselling Setting.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFHRL-TP-83-16	2. GOVT ACCESSION NO. AD-A126554	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Proceedings: 24th ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIATION		5. TYPE OF REPORT & PERIOD COVERED FINAL
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) MTA PUBLICATIONS COMMITTEE		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory(AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE 25 March 1983
		13. NUMBER OF PAGES 750
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

Proceedings of the 24th Annual Conference of the Military Testing Association, San Antonio, Texas, 1-5 November 1982

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

accession quality
all-volunteer force
aptitude
Armed Services Vocational Aptitude Battery (ASVAB), attrition
career education
classification (SEE REVERSE)
combat soldier
Comprehensive Occupational Data Analysis Programs (CODAP)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Military Testing Association Conference was held 1-5 November 1982 at the El Tropicano Hotel, San Antonio, Texas. It was coordinated jointly by the Air Force Human Resources Laboratory and the USAF Occupational Measurement Center. Independent presentations were made by members of the Department of Defense, United States Coast Guard, related Defense contractors, and representatives of Allied nations. The text of each presentation has been included in this report.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

computer assisted training
computer base instruction
counseling
criterion-referenced tests
data display
enlistment motivation
enlistment standards
fatigue
feedback
goal setting
individual productivity
Instructional Systems Development (ISD)
interrater reliability
job analysis
job description
job difficulty
job reading
job satisfaction
leadership
maintenance performance
management information system
models
motivation
Occupational Research Data Bank (ORDB)
occupational analysis
officer surveys
organizational assessment
organizational productivity
performance
performance evaluation
personality traits
personnel measurement
physical fitness
practice effects
problem solving
psychological screening
quality circles
rating policies
rating scales
readiness evaluation
recruiting
retraining
selection
simulation
socialization
soft skills analysis
stress
systems design
tactical symbology
tactics training
task analysis
task criticality
task inventories
test equating
testing
trainability
training
training evaluation
turnover
validation
vocational interests
women

AFHRL-TP-83-16 March 1983

PROCEEDINGS

24TH ANNUAL CONFERENCE

of the

MILITARY TESTING ASSOCIATION

coordinated jointly by

AIR FORCE HUMAN RESOURCES LABORATORY

and

USAF OCCUPATIONAL MEASUREMENT CENTER

A

San Antonio, Texas

1 - 5 November 1982

FOREWORD

As described in the Military Testing Association (MTA) by-laws the purpose of the organization is to assemble representatives of the various armed services of the United States and other nations that may request, to discuss and exchange ideas and to review and study research activities of associated organizations engaged in military personnel assessment. Further goals are to foster improved or new techniques and procedures for behavioral measurement, occupational & manpower analysis, simulation models, training programs, selection methodology and survey systems; to promote cooperation in the exchange of assessment procedures, techniques and instruments; and to promote assessment of military personnel as a scientific adjunct to modern military personnel management within the military and professional communities.

In 1982 primary affiliations of MTA included 12 armed services agencies, with associated governmental, educational, industrial, and private organizations engaged in activities that parallel the previously described purposes. The primary agencies were the the US Army Research Institute, US Naval Education and Training Program Development Center, US Navy Personnel Research and Development Center, US Coast Guard Institute, US Air Force Occupational Measurement Center, US Air Force Human Resources Laboratory, Royal Australian Air Force, Belgian Armed Forces Psychological Research Section, Canadian Forces Directorate of Military Structures, Canadian Forces Personnel Applied Research Unit, and Federal Republic of Germany Ministry of Defense. Also in 1982, the US Selective Service System Analysis and Evaluation Division was approved as a member agency, and it is anticipated that the Israeli Defense Forces will request membership.

The 24th Annual Conference of MTA was jointly coordinated by the Air Force Human Resources Laboratory and the USAF Occupational Measurement Center at the El Tropicano Hotel, San Antonio, Texas. The conference program began on 1 November 1982 with introductions, keynote address, and general session, and continued through 5 Nov 1982. Paper sessions and panels/symposia began on 2 Nov 82, with three simultaneous tracks of presentations. Various special interest/publication review groups and MTA steering committee meetings were held during the conference.

These proceedings document presentations made during 12 panels/symposia, 29 paper sessions, and general session. The presentations and discussions for which manuscripts or documentation were received and included represent a wide range of topics, issues, problems, activities, and research from the business, educational, governmental and military communities, both foreign & domestic. The papers reflect the opinions of their authors and are not to be construed as the official policy of any institution, government, or armed-service agency.

The 25th Annual Conference of the MTA will be coordinated by the Naval Education and Training Program Development Center. The 25th Anniversary Conference will be held during the week of 23 Oct 1983 at the Convention Center, Gulf Shores, Alabama.

WILLIAM C. DEBOE
Colonel, USAF
Chairman, MTA Conference

MTA 24th Annual Conference Staff

AFHRL and USAFOMC wish to thank the many people who have contributed to the planning and coordination of this 24th Annual Conference. The following is a list of committee chair persons/members who assisted in conducting the Conference:

<u>MTA Presidents</u>

Colonel Ronald W. Terry, USAF, Commander, AFHRL
Colonel Paul T. Ringenbach, USAF, Commander, USAFOMC

<u>Conference Chairman</u>

Colonel William C. DeBoe
Director, Applications and Liaison Office, AFHRL

<u>Committee Chair Persons/Members</u>

<u>Conference Coordinator</u>	
Dr. Joe T. Hazel	AFHRL
<u>Program</u>	
Dr. Walt Driskill	USAFOMC
Lt Colonel Jim Mitchell	USAFOMC
Mr. Jay Tartell	USAFOMC
Mr. Jim Keeth	USAFOMC
Lieutenant Eric Duncan	HQ AFSC
Dr. Linda Aslett	USAFOMC
<u>Membership/Registration</u>	
Ms. Joyce Giorgia	AFHRL
Ms. Cecilia Knutson	AFHRL
Ms. Barbara Arispe	AFHRL
<u>Publications - Proceedings</u>	
Ms Nancy Perrigo	AFHRL
Ms. Joyce Giorgia	AFHRL
<u>Facilities/Logistics</u>	
Captain Jim Kimbell	AFHRL
Mr. Phil Telisak	AFHRL
Senior Airman Ed Heugel	AFHRL
<u>Banquet</u>	
Ms. Joyce Giorgia	AFHRL
<u>Hospitality</u>	
Ms. Sharon Garcia	AFHRL
<u>Finance</u>	
Mr. Joe Hayden	AFHRL

TABLE OF CONTENTS

Page

Page

Page

NOTICE

Papers which were presented as part of SYMPOSIA/PANEL DISCUSSIONS are to be found under the title of the session in which they were presented. Sessions appear in the order in which they were listed in the program, and an overall abstract for each session is provided. Papers within the sessions are arranged alphabetically by principal author. We regret that a number of papers presented in the sessions are not included; however, we published all session papers received.

PAPER PRESENTATIONS are arranged alphabetically by principal author. The following rules were applied in determining which papers to publish: (a) paper must be no more than about six pages in length, and (b) author/presenter must have paid the MTA registration fee. An abstract was published in lieu of a paper in those few instances where the paper was not submitted for publication.

The Table of Contents at the front of the Proceedings and the Index of Authors in the back should be useful in locating specific papers of interest.

No attempt has been made to edit papers. The views expressed in them are those of their authors, and not necessarily those of the Military Testing Association or the organizations which the authors represent. Correctness is the sole responsibility of the authors.

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

RONALD W. TERRY, Colonel, USAF
Commander

OPENING SESSION OF THE 24TH ANNUAL
MILITARY TESTING ASSOCIATION CONFERENCE

1 November 1982

Welcome - Colonel Ronald W. Terry, Commander, Air Force Human Resources Laboratory, officially opened the 24th Annual MTA Conference at 1300 hours, 1 November 1982. On behalf of AFHRL and the USAF Occupational Measurement Center, cohosts for this year's conference, Colonel Terry welcomed all participants to San Antonio and to the conference, and particularly noted the participation of military representatives of Australia, Belgium, Canada, Israel, the United Kingdom, and West Germany. He briefly reviewed the importance of Human Resources research and applications to the US Air Force operational mission and noted a number of research thrusts currently underway or programmed which will impact on the Air Force operational readiness. Colonel Terry also expressed the hope that the conference could help the cause of operational readiness through a frank exchange of ideas and research results among all participants.

Introduction - Other key conference officials on the podium included: Colonel Paul T. Ringenbach, Commander of the USAF Occupational Measurement Center and cohost for this conference; Colonel William C. DeBoe, Director of the AFHRL Applications and Liaison Office and MTA Conference Chairman; and Dr. Walter E. Driskill, Chief, Occupational Analysis Program, USAFOMC, and Chairman of the MTA Program Committee. Colonel Ringenbach then introduced the Keynote Speaker for the Conference, Major General Spence M. Armstrong, Commander, Air Force Military Training Center, Lackland AFB, Texas. Colonel Ringenbach briefly traced General Armstrong's career from early rated assignments to a graduate engineering degree program, to the Air Staff at HQ USAF. In 1980-1981, General Armstrong was the Deputy Chief of Staff for Technical Training, Headquarters Air Training Command, at Randolph AFB, Texas, where he was responsible for all Air Force technical training, mobile training teams, field training detachments, career development courses promotion test development, and occupational analysis. In mid-1981, upon his promotion to Major General, he was reassigned to command AFMTC at Lackland, where all Air Force basic military training is conducted.

Keynote - Major General Armstrong welcomed all conference participants to San Antonio on behalf of the Commander, ATC, and the USAF. He discussed "in agricultural terms" his experiences with the Air Force program development process, including its problems in terms of funding, personnel resources, recruiting, and training. General Armstrong described some of the key areas of the programming/planning cycle and how system development and procurement actions drive future manpower and training needs. These

programs must then be translated into specific recruiting goals for individuals possessing the capability to learn highly technical maintenance or operational skills. Training programs must be developed or modified to meet the changing technical and operational systems. At many of the key phases of this process, research and applications personnel are making very important decisions in terms of human resources requirements and development programs. It is imperative for personnel in the scientific community who are involved in such decisionmaking to keep in touch with the operational Air Force. Their high degree of technical involvement in their own areas often makes it difficult to communicate what they are doing to others, and thus their impact on systems decisions can sometimes be limited. To be fully effective, such highly technical specialists must ensure that their work relates to the real needs of Air Force operational programs. Just as important is for those working in research and analysis to assist Air Force operators to gain an understanding of how to apply the valuable research developed. Research is valuable only if it is used in making important Air Force decisions.

General Session - The conference was reconvened at 1500 hours by Colonel William C. DeBoe, MTA Chairman. After administrative announcements, the general session was devoted to presentations by representatives of several Allied Nations: Commandant Arnold C. Bõhrer, Belgian Armed Forces; Dr. Heinz-Jurgen Ebenrett, Federal Armed Forces (West Germany); and Captain Harold Mendes, Canadian Forces (Summaries of these presentations are included in the Papers section of this volume).

SYMPOSIA/PANEL DISCUSSIONS

SOFT SKILLS ANALYSIS

Chair: Eva L. Baker

Participants described their work in the areas of soft skills analysis and measurement, and identified applications for improving practice in this difficult area. Members represented academic, civilian, and military researchers, as well as practitioners.

AD P000807

SOFT SKILL ANALYSIS:

TWO PROPOSED METHODS FOR ANALYSIS

PREPARED FOR

MILITARY TESTING ASSOCIATION

BY

MAJOR RONALD W. TARR

US ARMY, TRAINING DEVELOPMENTS INSTITUTE
Fort Monroe, VA 23651

ABSTRACT

In 1975 the US Army adopted a state-of-the-art systems model for the development of training. The focus of this model was on the greatest training requirement facing the Army Training and Doctrine Command(TRADOC), which was and still is, the initial entry training of enlisted soldiers. As such, the type of jobs these junior enlisted soldiers do are made up predominantly of procedural tasks, thus the highly detailed ISD procedures were developed to work primarily on such tasks. By 1976, TRADOC realized that the smaller, yet highly critical behaviors of senior NCOs and officers training requirements had to be addressed. From a large meeting held in 1976(Soft Skill Symposium) began the lengthy process that has led to the two analysis approaches recommneded by this author. The first approach focuses on very complex tasks, and is called the Extended Task Analysis Procedures(ETAP). This process was initially concieved by the original authors of ISD, and is intended as an extension of those procedures. The second approach was developed by the author as a result of several years of study and examination of the problem as well as a variety of other solutions. This approach, Complex Skills Analysis, looks at the total job requirements of an encumbent, and attempts to sort out and identify what "soft skills" exist, how they fit and interact with the more discrete activities and tasks and then to carefully examine these for their requirements and quality. This approach and the ETAP model plus a recent effort conducted under contract, together make up a total package for the analysis and documentation of these highly complex behaviors previously rather ignored under ISD.

The attached is an extract from the package and is a draft of the Complex Skills process. Copies of the complete package can be acquired by contacting the author.

The inclosed comments are those of the author and do not necessarily represent the views or opinions of the United States Army.

A PROPOSED PROCESS FOR "SOFT SKILLS" ANALYSIS

The Complex Skills Analysis, stated above, is a general process that is intended to guide the analyst through examination of these behaviors. It is not intended to be a checklist or lockstep procedure but rather a way of focusing the analyst's efforts. The process has been developed by the author via several years of dealing with the analysis of these complex skills and the synthesis of numerous existing techniques. It is in no way an approved or foolproof solution, but it will be a recommended process for use by the TRADOC community. Remember, doing this type of analysis is one of the best examples of one of these complex skills!

Step I: Identify Candidate Complex Skills

Using an SME who is familiar with the complete field to be studied, develop an initial list of skills that are fairly obvious as part of the job. For instance, it is fairly clear that if the job calls for supervision and leadership activities that the skills will, as a minimum, include communications and interpersonal skills. This is not a desk top or opinion oriented analysis but merely a common sense point to start. The substeps below will guide this behavior.

Substep A: Taking no notes, and after establishing your purpose with the SME, ask him to describe the job or position in very general terms.

Explain that you must be familiar with the limits of the job, its parameters, dimensions, and complexities, so you can understand the context of the behaviors involved. As much as possible, assist the SME if focusing on behavior and performance are major pieces of the job.

Substep B: With the SME, review the results of any Job Analysis that have been previously conducted.

Remember the importance of a Job Analysis to this process. At this time, you should identify what kind and how useful (if at all) the Job Analysis is. In some cases none will exist. This could be because it is a new job or position or conflicting requirements simply precluded it being done.

////NOTE: DO NOT BECOME JUDGMENTAL IF A JOB ANALYSIS HAS NOT BEEN DONE////
This could serve only to block your effort and would not help.

What you are looking for in the Job Analysis is more concrete information about the job. For instance, who is currently filling (type of folks), information about conditions, equipment involved, relative size of job, and what procedural tasks have been identified as making up the job. These will be very helpful later in the process.

If no Job Analysis is available, then a careful examination with documentation should be accomplished. The interaction of the complex skills throughout the entire job is an integral portion of the skills themselves. This may be something that will delay the process but is essential. It may be possible to have it conducted at the same time by other members of your office.

Regardless, the output of this substep must be a clear definition of the total job which is shared by the SME and the analyst and either the Job Analysis documentation or the analyst's substitute. (The analyst's substitute should not be considered a replacement for a Job Analysis for other purposes.)

Substep C: Now taking notes, begin interviewing the SME about relevant major activities about the job or position.

This is not to duplicate the Job Analysis but rather to focus the SME on the key interactions of the activities and to provide some important context relationships for the analyst. Again, focusing the SME on actions should assist him and prevent too much philosophizing or "war stories."

////////////NOTE: DO NOT TURN OF THE SME BY BEING TASK ORIENTED////////////
This activity is to provide flavoring and interrelationships.

/////////////NOTE: DO NOT LET THE SME BECOME HUNG UP ON DETAILS/////////////
Be sure he knows you will get details later and want generalities.

Substep D: Using the results of the above steps, develop a list of complex skills that make sense to you and the SME.

This list should be a joint effort in which you and the SME are deriving statements of generic behavior from your focus on the total performance involved but at the job level. This list might look like the following:

Conduct Inspections

Delegating responsiblity

Motivating subordinates

Effectively communicate

Supervise subordinates

Substep E: Compare each item on the list against the job performances to cross check that the skills match and are relevant candidates.

This is a final check that should consist of each separate skill being examined in its relationship to the total job. This could be compared to a Murder Board for selecting critical tasks in which each task is examined for its relationship to the job.

/////NOTE: THIS IS BY NO MEANS THE END OF THE LISTING OR IDENTIFICATION/////

STEP II: Establish the context of the behavior.

Substep A: Ask the SME to describe how the behavior fits into the overall job--in detail but at job/duty level.

Substep B: Ask the SME how the behavior/skill interacts with other skills; do they support each other; are they dependent on others; do they cue each other, etc.

Substep C: Ask the SME what tasks are directly related to the "SS;" does the "SS" run through several tasks; is it an integral part of a large complex task; is the "SS" a transfer task complete (counseling)? (Here the JA will help.)

STEP III: Select one complex task or scenario as basis for analysis.

Substep A: If the SME says that a sequence exists (similar to a procedure), ask him to select one that is typical or representative to use as the "base piece" for the analysis.

Substep B: If sequence is not readily identifiable, ask the SME to help you come up with a simulated scenario that would require "SS" application. The scenario should depict a typical setting, be of sufficient realistic detail to be valid, include the things that would initiate "SS," and require it to be properly applied.

STEP IV: Conduct initial analysis interview.

Substep A: Ask the SME to explain, based on the selected procedure or scenario, what happens; that you will stop him to ask why he did or didn't do something (decision points) along the way. Start by asking him what initiates the "SS," how he knows to apply the skill, and what he considers prior to actually starting. Don't get him into an exception loop or let him get bogged down. Further examination will happen later. The process (decomposing) will happen shortly but not yet. He should describe the behavior at the same level as the preceding activity, but now you will make notes and looking/listening for specifics, key phrases, terms, decisions, rules, cautions, and etc., or anything that crystalizes the concept.

STEP V: Repeat the above for each step or piece of the scenario. This will provide you with about four to eight steps/pieces. If you have less than that, they are probably too big; more, and they are probably too small. However, before you try reconstructing them, have the SME review them and the total "SS." It may be that this behavior is too big or too little. As much as possible, one must try to standardize the size of the "SS" being analyzed. This is important during analysis as well as for the follow-on designer/developers. Don't force it. "SS" vary in size and scope, but they also intertwine pretty tightly sometimes and are hard to sort out. Assist the SME by helping him focus on the main outcome or product of the "SS"; this can help him strip away other actions or behaviors surrounding it that don't affect it. Don't strip away a dependent or supporting skill or piece in your zeal to be analytical. Size is important but is nothing compared to validity. When you are satisified the "SS" is the right size, but you are still under four or over eight, review each step/piece with the SME to see if they can be split or lumped together. Some adjustment and fine tuning can always be done when you decompose the step/pieces. The diagram below shows what you might have at this point of the analysis.

"SS" DESCRIPTION

Label #1	Label #2	Label #3	Label #4	Label #5
Detailed Description of #1	Details of #2	Details of #3	Details of #4	Details of #5

STEP VI: Decomposition - Level 2

During this phase, you will be breaking down or decomposing each Level 1 piece into actions, rules, decisions, and etc. It will also allow you to identify task activities from generic skills. The mixture of tasks and generic skills can be likened to a mosaic, with the tasks being the tiles and the generic skills the cement that acts on many of the tiles and binds them into a total job behavior.

As in any analysis, try to have the SME describe what he does, thinks, and etc., in behavioral terms. Use verbs that are clear and indicate peformance. The decomposition process consists simply of having the SME do the same thing for each major step as he did for the task.

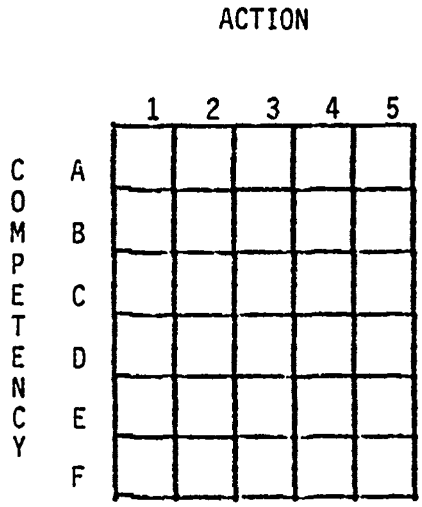
It is very important that you clearly identify action steps (observable performance) from decision steps (something happening internal; rules applied). After the steps get broken down and labeled, you will begin sorting out performance from the support skills and knowledges and evidence/

indications of generic skills. NOTE: Probably one of the most basic characteristics of soft skills that makes analysis difficult is that they are rarely, if ever, by themselves. They run through all types of "hard skills" and mix among themselves. They have no clear beginning or end; they are used in varying amounts by different people in different situations.

In analyzing them, you must approach them by carefully stripping away the procedural aspects of a total behavior and then examine what is left. By observing behavior, we can note what is actually happening, identify what caused it, why it was done the way it was, and what decisions were made (also what specific knowledge).

One of the beauties of competency is that they can be used for other than job situations and that gives the student ample practice and reinforcement opportunities. For training this is great--for analysis it must muddies up the water. This characteristic of competencies being integrated is a powerful key to the whole issue. It requires the analysis be conducted differently, especially the documentation of results, and more importantly that training be designed and developed differently. This means that the training itself should be integrated, if the proper performance outcomes are to be achieved. For instance, if one is developing task oriented training, the instruction consisted of usually three phases: task presentation, task practice and feedback, and then evaluation. For competency based training, there should be at least five phases: competency presentation, integration with task presentation, structured practice, and feedback (by the numbers) in which the interaction of competence and task are clearly indicated 2nd P-F on different task (where both task performance and competency application are tested). What this does is clearly demonstrate the generalizability of the competency across several task; plus, it greatly enhances the student's ability to generalize it to other new situations. Since the overall goal of a leader is to take appropriate action when faced with a situation based on his training, experiences, and doctrine when this transferability is absolutely essential. Many times in the past we have seen task oriented training conducted and people surprised when it didn't generalize to a new job situation. Straightforward facts and procedures are so specific that they rarely generalize. This is probably because the details peculiar to the procedures are taught (and usually emphasized) in the same way as the parts that might generalize. For example, if you are taught that X is a generator on engine Y, then everytime you see engine Y you can easily identify generator X. But, if I tell you all engines have a generator, and on engine Y it is located at X and on engine Z, it is a U then you would be more likely to learn the concept of generators which would generalize to engine A through T; that all have generators but not in the same place or looking exactly the same. The concept of generators is not important unless we expect the student to generalize the concept to other new situations. If he is a 63W, Wheeled Vehicle Ordnance Mechanic, responsible for 68 vehicles, then this becomes pretty important. And it becomes really good when he is called upon to work on an M60 tank because the tank man is wounded.

A way of graphically showing some of this stuff is by use of a Matrix in which we list actions/tasks across the top and generic skills/competencies down the side (below).



This shows how: (1) several competencies impact on a single task, as well as, (2) how a single competency spreads over several tasks. Unfortunately, this is very difficult because they are fairly separate but above we said they interact. The results of this interaction could be called a performance outcome and could be a description of what the complex behavior means, as an example of success field performance (case study).

AD P000808

ON DISCRETE DISCRIMINANT ANALYSIS

Rand R. Wilcox

Department of Psychology
University of Southern California
and
Center for the Study of Evaluation
University of California, Los Angeles

When predicting group membership, success on an external criterion, mastery in a particular subject area, etc., there are of course many discriminant analysis procedures that might be applied. In many cases the techniques considered are a function of the type of data that is available. The goal in this paper is to (1) discuss some of the author's recent work that is relevant to discriminant analysis, (2) comment on recent related investigations by other investigators, and (3) suggest directions for future research.

Discrete Discriminant Analysis with Binary Random Variables

First consider the situation where for each subject there are k+1 measures, say $x_1,\ldots,x_k,y$ where x_i $(i = 1,\ldots,k)$ and y take on the values 0 or 1. As is customary, a value of 1 might mean the presence of some characteristic, or success on some task. The random variable y is some external criterion of interest such as a subject being judged to work well within some group or team of individuals. It is assumed that the values of $x_1,\ldots,x_n,y$ have been observed for N individuals, and that for future subjects the goal is to predict y from the observed values of $x_1,\ldots,x_k$. Many solutions to this problem have been posed (e.g., Ott and Kronmal, 1976; Dillon and Goldstein, 1978; Aitchison and Aitkin, 1976). Perhaps the best known solution is Fisher's linear discriminate function, but for the situation at hand, it is known to be unsatisfactory (e.g., Goldstein and Dillon, 1978).

Most other solutions to predicting y from $x_1,\ldots,x_k$ are based on estimates of the joint probability function of $x_1,\ldots,x_k,y$, say $f(x,y)$. Among the N subjects for whom there exists information about both $x = (x_1,\ldots,x_k)$ and y, let N_{xy} be the number of subjects with an observed x and y. Then $\hat{f}(x,y) = N_{xy}/N$ is the usual unbiased estimate of $f(x,y)$. If $\alpha = Pr(y=1)$, then the optimal rule for predicting y, given x, is to predict y = 1 if

$$\alpha f(x|y=1) \geq (1 - \alpha)\, f(x|y=0); \tag{1}$$

otherwise predict y = 0 (Anderson, 1958, p. 130; cf. Copas, 1974). Of course α is usually unknown, but it can be estimated with the proportion of subjects having y = 1, and this together with $\hat{f}(x,y)$ yields an estimate of the optimal rule given by (1).

Several alternative estimates of (1) have been proposed, four of which were empirically compared by Wilcox (1980). The procedure that performed

best was one proposed by Aitchison and Aitkin (1976) where f(x,y) was estimated with

$$\hat{f}(x,y) = N^{-1} \sum_{i=1}^{N} K(x_i, y_i, \lambda) \tag{2}$$

where the vector x_i and scalar y_i are the values of x and y for the $i^{\underline{th}}$ subject,

$$K(x_i, y_i, \lambda) = \lambda^{k+1-d} (1-\lambda)^{d}$$

$d = d(x_i, y_i; x, y)$ is the number of components in disagreement between the vectors (x_i, y_i) and (x,y), and λ is an unknown parameter that is estimated from the data. The procedure suggested by Dillon and Goldstein (1978) gave the poorest results, even compared to using $f(x,y) = N_{xy}/N$. The other procedure considered was one proposed by Ott and Kronmal (1976).

Despite the advantages of (2) listed by Aitchison and Aitkin (1976), some caution must be used. In particular, Hall (1981) points out that Aitchison and Aitkin's estimate of λ can behave erratically, even with large samples, and that it is strongly influenced by the presence of empty cells, or cells having only one observation. Hall goes on to discuss ways of correcting this problem (cf. Wang and Van Ryzin, 1981; Bowman, 1980).

Comments on Monte Carlo Studies of Discrete Discriminant Analysis Procedures

In addition to the results in Wilcox (1980), there are Monte Carlo Studies that also indicate that it is generally possible to improve upon $\hat{f}(x,y) = N_{xy}/N$ in terms of predicting y from x. Most of these studies generate observations using a two-term approximation of the multinomial distribution proposed by Bahadur (1961). An important question is whether this approximation works well when k is large, for example, when k = 9. Put another way, the probability functions used to generate observations are assumed to have a particular structure, but the extent to which these structures approximate real data sets was never made clear.

From Wilcox (1982a, 1982b) it appears that a two-term Bahadur approximation of multinomial distributions generally works well for k = 4 and possibly for k = 5, but for k = 9 this is not the case. A three term approximation was also tried (but never published), and unfortunately it seems to give little improvement. The implication is that certain multinomial distributions are difficult to approximate using the procedure in these Monte Carlo Studies, and so there is some doubt about whether these studies generalize much beyond k = 4. Ott and Kronmal (1976) used a representation of binary data proposed by Good (1963), but the same concern seems to apply.

Currently it can be said that when k is relatively small, it is frequently -- but not always -- possible to improve upon $f(x,y) = N_{xy}/N$ when estimating (1). A reasonable speculation would be that this result will hold for larger values of k, but this has not been established.

Determining Passing Scores

In some situations it may be desirable to determine a passing score for predicting success or failure on some external criterion (Huynh, 1976). For instance, a subject might be given a test, the possible values of which are $w = 0,\ldots,n$. The goal might be to find a w_0 with the idea that if $w \geq w_0$, predict $y = 1$; otherwise predict $y = 0$.

A method for determining the better of two passing scores was proposed by Wilcox (1979). The situation can be breifly summarized as follows. Consider two passing scores, say w_{01} and w_{02}. The six possible outcomes and their associated probabilities are given in Table 1. Thus, t_{11} is the probability of having $w \geq w_{02}$ and $y = 1$. the probability of incorrectly predicting y using w_{02} is $t_{10} + t_{21} + t_{31}$, and the corresponding probability for w_{01} is $t_{10} + t_{20} + t_{31}$. Thus choosing the optimal passing score reduces to determining whether t_{21} is less of greater than t_{20}.

Table 1

Probabilities Associated with Two Passing Scores

	$y = 1$	$y = 0$
$w \geq w_{02}$	t_{11}	t_{10}
$w_{02} > w \geq w_{01}$	t_{21}	t_{20}
$w \quad w_{01}$	t_{31}	t_{30}

For convenience let $p_{11} = t_{11} + t_{31}$, $p_{00} = t_{10} + t_{30}$, $p_{10} = p_{21}$ and $p_{01} = t_{20}$. Also let $\hat{p}_{ij}$ be the usual unbiased estimate of p_{ij} ($i = 0,1$; $j = 0,1$). Then the goal is to choose N so that

$$\Pr(\hat{p}_{10} \geq \hat{p}_{01}) \geq p^* \tag{3}$$

whenever $p_{10} - p_{01} \geq \delta^*$, where $\delta^* > 0$ is chosen by the experimenter; the constant δ^* is the smallest difference between p_{10} and p_{01} that an investigator is concerned about. If $\hat{p}_{10} = \hat{p}_{01}$, one of the two passing scores is chosen at random. Wilcox's results indicate that a large N might be required to satisfy (3). When considering several passing scores the problem gets worse.

Let $p_1 = p_{11} + p_{10}$ and $p_2 = p_{11} + p_{01}$ in which case determining whether p_{10} is larger than p_{01} is the same as determining whether p_1 is larger than p_2. Thus, results in Tamhane (1980) are the same as those in Wilcox (1979)

except that Tamhane's includes the additional requirement that (3) be satisfied whenever

$$p_1 + p_2 \leq \gamma^* \tag{4}$$

For $\gamma^* = 1$ the situation reduces to the one considered by Wilcox (1979), and for $\gamma^* < 1$ a smaller N is needed to satisfy (3). Hence, if a γ^* can be specified, having to use a large sample of subjects to determine the optimal passing score might be avoided (cf. Lam and Mehra, 1981).

Further Comments on Estimating the Optimal Decision Rule

Consider any measure, say w, and again suppose the goal is to predict y given w. If the distribution of w given y can be estimated, this yields an estimate of the optimal rule given by (1) except that $f(x|y)$ is replaced by $f(w|y)$, the probability density function of w given y. A common assumption is that $f(w|y)$ is normal, but for many situations it might be more realistic to assume unimodality, but allow for the possibility that the distribution is skewed. Let a and b be the minimum and maximum possible possible values of w. When a and b are known it might also be useful to take this information into account.

If unimodality can be assumed, then a very good approximation of $z = (w-a)/(b-a)$ might be possible using a beta distribution (Springer, 1979; Weiler, 1965). Smith et al. (1981) found such an approximation useful, and Wilcox (in press) indicates that this approach seems to improve upon the usual chi-square approximation of the X^2 statistic used to test for equiprobable cells in a multinomial distribution.

The estimate of $f(z|y)$ is as follows. Suppose that for y = 1, the observed z values are $z_1,\ldots,z_N$. Let $\hat{\mu}$ and $\hat{\sigma}^2$ be the usual estimates of the mean and variance of z. The beta distribution is given by

$$f(t) = \frac{\Gamma(r+s)}{\Gamma(r)\,\Gamma(s)}\, t^{r-1}\,(1-t)^{s-1} \tag{5}$$

where $r > 0$ and $s > 0$ are unknown parameters, and Γ is the usual gamma function. The estimates of r and s are

$$r = \hat{\mu}^2\,(1-\hat{\mu})/\hat{\sigma}^2 - \hat{\mu}$$

and

$$s = \hat{\mu}\,(1-\hat{\mu})^2/\hat{\sigma}^2 + \hat{\mu} - 1$$

Of course the estimate of $f(z|y)$ for y = 0 is calculated in the same manner.

REFERENCES

Aitchison, J. & Aitken, C.G.G. Multivariate binary discrimination by the kernel method. Biometrika, 1976, 63, 413-420.

Anderson, T.W. An introduction to mutlivariate statistical analysis. New York: John Wiley & Sons, 1958.

Bahadur, R.R. A representation of the joint distribution of responses to n dichotomous items. In H. Solomon (Ed.) Studies in item analysis and prediction. Stanford: Stanford University Press, 1961.

Bowman, A.W. A note on consistency of the kernel method for the analysis of categorical data. Biometrika, 1980, 67, 682-684.

Copas, J.B. On symmetric compound decision rules for dichotomies. Annals of Statistics, 1974, 2, 199-204.

Dillon, W.R., & Goldstein, M. On the performance of some multinomial classification rules. Journal of the American Statistical Association 1978, 73, 305-313.

Goldstein, M. & Dillon W.R. Discrete discriminant analysis. New York: John Wiley & Sons, 1978.

Good, I.J. Maximum entropy for hypothesis formulation, especially for multidimensional contingency table. The Annals of Mathematical Statistics, 1963, 34, 911-934.

Hall, P. On nonparametric multivariate binary discrimination. Biometrika, 1981, 68, 287-294.

Hunyh, H. Statistical consideration of mastery scores. Psychometrika, 1976, 41, 65-78.

Lam, K. & Mehra, K. A note on some results of Tamhane. Journal of the Royal Statistical Society, 1981, 43, 368-369.

Moore II, D.H. Evaluation of five discrimination procedures for binary variables. Journal of the American Statistical Association, 1973, 68, 399-404.

Ott, J. & Kronmal, R.A. Some classifcation procedures for multivariate binary data using orthogonal functions. Journal of the American Statistical Association, 1976, 71, 391-399.

Smith, P.J., Rae, D.S., Manderscheid, R.W. & Silbergeld, S. Approximating the moments and distribution of the likelihood ratio statistic for multinomial goodness of fit. Journal of the American Statistical Association, 1981, 76, 737-740.

Springer, M.D. The algebra of random variables. New York: John Wiley & Sons, 1979.

Tamhane, A.C. Selecting the better Bernoulli treatment using a matched samples design. Journal of the Royal Statistical Society, 1980, series B, 24, 26-30.

Van Ryzin, J. & Wang, M.C. Two methods for smooth estimation of discrete distributions. Communications in Statistics, 1978, A7, 211-228.

Weiler, H. The use of incomplete beta functions for prior distributions in binomial sampling. Technometrics, 1965, 7, 335-347.

Wang, M. & Van Ryzin, J. A class of smooth estimators for discrete distributions. Biometrika, 1981, 68, 301-309.

Wilcox, R.R. A lower bound to the probability of choosing the optimal passing score for a mastery test when there is an external criterion. Psychometrika, 1979, 44, 245-249.

Wilcox, R.R. An empirical comparison of four discrete discriminant analysis procedures. Educational and Psychological Measurement, 1980, 40, 981-986.

Wilcox, R.R. Approximating multivariate distributions. Center for the Study of Evaluation. University of California, Los Angeles, 1982a.

Wilcox, R.R. An approximation of the k out of n reliability of a test, and a scoring procedure for determining which items an examinee knows. Submitted for publication, 1982b.

Wilcox, R.R. A comment on approximating the X^2 distribution in the equiprobable case. Communications in Statistics--Simulation and Computation, in press.

INDIVIDUAL AND ORGANIZATIONAL PRODUCTIVITY: RESEARCH AND ISSUES

Chair: Michael Matthews

Issues related to productivity enhancement and productivity measurement in the military environment were discussed. Empirical studies showing the relationship between quality circles and productivity, informational feedback and productivity, and job satisfaction and productivity were presented. In addition, a comprehensive methodology designed to generate objective criteria of organizational productivity was described. Finally, the use of objective measures of productivity measurement in management consulting was discussed. The research presented should be of interest to personnel involved in productivity management, as well as to other researchers in the field.

AD P000809

Predicting Job Satisfaction and Job Performance

Gene A. Berry and Michael D. Matthews
Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 7823

The Air Force is very concerned with obtaining the fullest possible utilization of its personnel resources. A critical part of that goal requires that incoming personnel be assigned to jobs that will optimally match their abilities and interests. In evaluating an individual's placement among potential assignments, present placement procedures rely primarily on the results of aptitude testing, job entry requirements, and needs of the service. An applicant's vocational preferences, with respect to available jobs, are typically assessed only on an informal basis during conversations with Air Force recruiting or occupational counseling personnel. Although some choice may be exercised on the part of the applicant during this process, decisions are sometimes made which are less than optimal. Additionally, since persons entering the service typically have little prior experience in the job market, and very little knowledge of the Air Force occupational system, they understandably have a difficult time relating personal likes and dislikes to the job choices available. However, the consequences of misclassification at the entry level can be very costly for both the individual and the Air Force.

To minimize the probability of job misclassification, an interest assessment instrument was developed. The Vocational Interest-Career Examination (VOICE) is an Air Force instrument designed to assess vocational interests among Air Force enlistees. Its development and validation are described by Alley and Matthews (1982). In addition to measuring vocational interest, research has shown that job satisfaction can be predicted by the VOICE (Alley, Wilbourn, & Berberich, 1976). Job satisfaction has been found to be related to fatigue, dissatisfaction with life, depression, psychosomatic illness, mental illness, drug and alcohol abuse, job performance, and coronary heart disease (Cf. Alley & Matthews, 1982). Perhaps an equally serious implication of personnel dissatisfaction, however, has to do with its influence on various forms of occupational withdrawal. Research has demonstrated quite consistently that personnel dissatisfied with their jobs are much more likely to be absent from their work (Waters & Roach, 1973) and to terminate their employment at a higher frequency than are satisfied workers (Mobley, Griffeth, Hand, & Meglino, 1979).

The diverse and serious implications of job dissatisfaction led the Air Force Human Resources Laboratory to initiate a study of the relationship between vocational interests among first-term enlisted accessions, as assessed by the VOICE, and later occupational outcomes. Preliminary results on the relationship between job satisfaction, as predicted by the VOICE, and turnover have been presented earlier by Matthews (1982) and Matthews and Berry (1982). While both non-attrition and reenlistment are extremely desirable and can probably be influenced through improved initial assignments, another behavior, job performance, is also important. Previous research has found relationships between job satisfaction and job performance (Cf. Seashore & Taber, 1975). The purpose of this paper is to describe preliminary findings summarizing the relationship between predicted job satisfaction, as assessed by the VOICE at time of enlistment in the Air Force, and subsequent performance on the job.

Method

Sample

The VOICE was administered to a sample of 3,782 1979 and 1980 Air Force enlistees during their first week of Basic Military Training. The subjects were tracked during their initial tour of duty, and ratings of their job performance were obtained during either their second year of active duty (1980 enlistees) or their third year of active duty (1979 enlistees). The subjects were typical of first-term Air Force enlistees in terms of racial composition, age, and educational level.

The VOICE

The VOICE consists of a 300-item vocational interest inventory requiring approximately 30 minutes to administer. Individual items are presented in booklet form and consist of occupational titles, work tasks, leisure time activities, and desired learning experiences. Respondents indicate relative preferences for each item in a standard like-indifferent-dislike (LID) format. Item responses were converted to two types of scales: (a) basic interest scales, and (b) occupational scales. The basic scales represent measures of general interest in various occupational and technical areas. They were constructed by grouping items of similar content into 18 independent sets covering a wide range of interests in the vocational and technical domain. The basic interest scales cover areas of Office Administration, Electronics, Heavy Construction, Science, Outdoors, Medical Service, Aesthetics, Mechanics, Food Service, Law Enforcement, Audiographics, Mathematics, Agriculture, Teacher/Counseling, Marksman, Craftsman, Drafting, and Automated Data Processing. All items within each scale are homogeneous in the sense that each was selected to measure the same underlying dimension. The Office Administration items, for example, measure interest in clerical, administrative, and business related activities.

The occupational scales were designed for use in evaluating job assignment alternatives. It has been found that certain patterns of basic interest scores predict job satisfaction in various Air Force job clusters (Alley et al., 1976). These clusters, 20 in number, represent an exhaustive categorization of Air Force job specialties. The VOICE occupational scales, therefore, provide a predicted job satisfaction score for each of these 20 job clusters. Consequently, if used operationally job placement personnel would be able to readily obtain a prediction of job satisfaction for any Air Force career field, by determining in which of the clusters that particular job falls. The occupational scales, while formulated from basic interests, provide direct estimates of job satisfaction for each career field in the set and can be used for making specific comparisons between alternative assignments (Alley et al., 1976). Predicted job satisfaction (PJS) scores range from 200 to 800, with a mean of 500 and standard deviation of 100. For a more thorough and technical discussion of the development of the VOICE and a description of the basic interest and occupational scales, their psychometric characteristics, and validity, see Alley and Matthews (1982).

Procedure

A Job Performance Questionnaire was sent to the immediate supervisor of each airman in the sample. Each supervisor was asked to respond to the following items, comparing the airman he/she supervised to others performing the same types of duties and possessing similar job experience:

> How well does this person understand the technical aspects of his or her job?
>
> How motivated does this person seem to be to do a good job?
>
> How well does this person perform assigned duties?
>
> How well does this person appear to be progressing toward performing in a supervisory role in his or her job?

Responses were made using a seven point scale ranging from "Very much above average" (7) to "Very much below average" (1). Questionnaires were returned by the supervisors of 798 male and 167 female 1979 enlistees, and 1,272 male and 512 female 1980 enlistees, for an overall return rate of 73 percent of the original sample of 3,782.

The VOICE predicted job satisfaction score corresponding to the career field in which each airman was assigned was determined. This information, along with Armed Forces Qualification Test (AFQT) scores, was used in predicting the criterion variable, rated work performance.

Results and Discussion

Responses to the four items on the Job Performance Questionnaire were totaled to provide an estimate of overall work performance. The main findings of the study are presented in Figure 1, which depicts rated work performance, collapsed across year of enlistment and gender as a function of predicted job satisfaction. Subjects assigned to career fields with associated low predicted job satisfaction had an overall work performance rating of 19, which, when divided by the number of items (four), shows an average item rating of 4.75, or "average" to "slightly above average" on the seven point Job Performance Questionnaire scale. Personnel assigned to jobs with intermediate levels of predicted job satisfaction had a total score of 20, or a mean of 5.0 which was "slightly above average" on the seven point scale. Personnel assigned to jobs in which they had high predicted job satisfaction had a total rating of 22, or an average rating of 5.5, which would be between "slightly above average" and "above average" on the seven point scale. A regression model with vectors for VOICE predicted job satisfaction scores and AFQT scores was developed to predict rated work performance. Analyses showed an R of .080 (F=8.78; df=2, 2,746; $p < .05$) for AFQT scores and VOICE scores combined. The AFQT alone had an R of .063 (F=11.13; df=1, 2,747; $p < .05$), and the VOICE alone had an R of .053 (F=7.90; df=1, 2,747; $p < .05$) with rated work performance.

One factor that would tend to limit the magnitude of the relationship between VOICE predicted job satisfaction scores and rated work performance is the fact that most (70%) of the 1,033 subjects for which rated job performance data were not obtained had attrited from the Air Force. It has been shown that

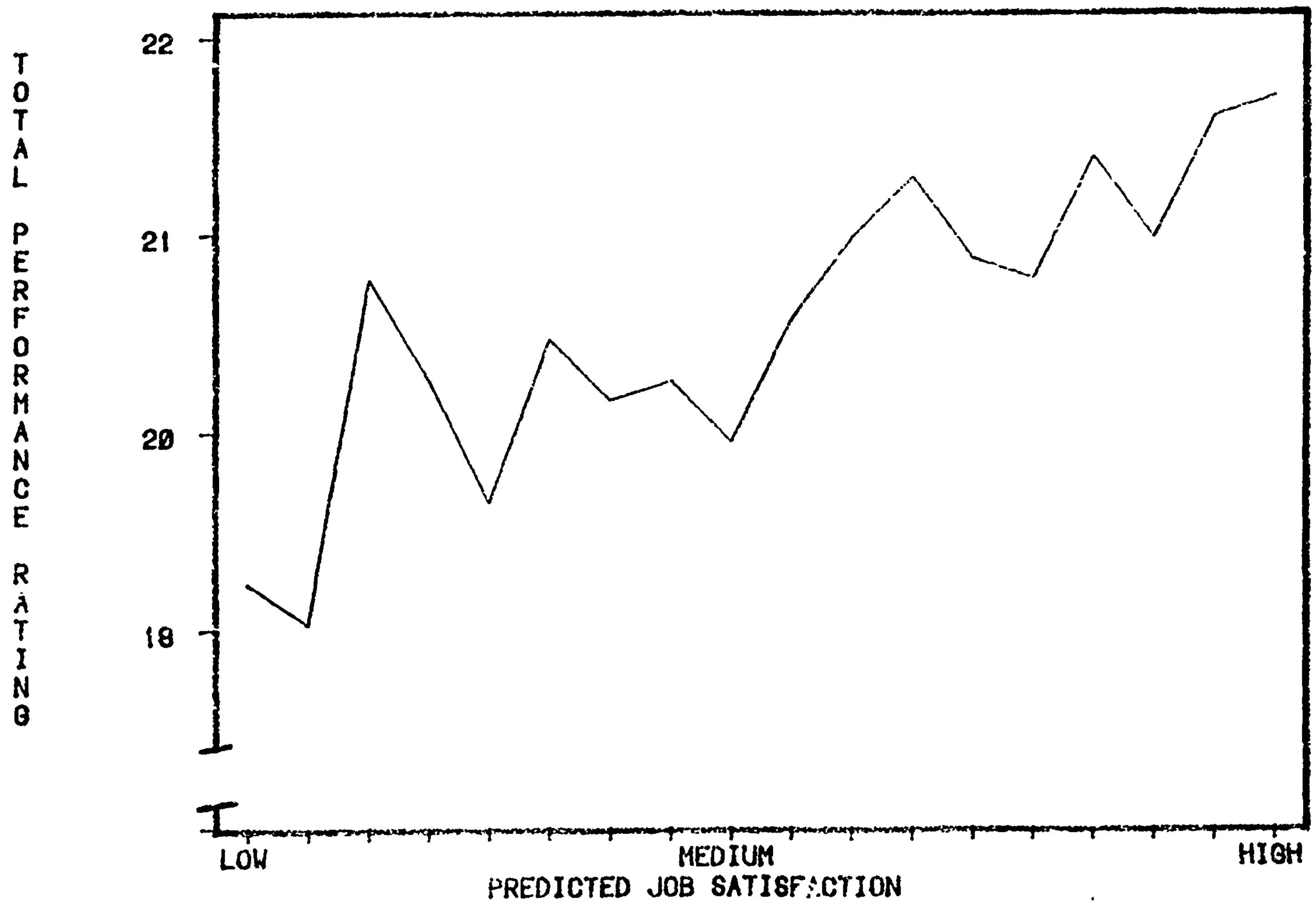


Figure 1. Actual job satisfaction as a function of predicted job satisfaction.

Air Force first-term attrition is related to VOICE predicted job satisfaction scores, with low predicted job satisfaction associated with high attrition rates (Matthews, 1982; Matthews & Berry, 1982). Moreover, many of these attritions were probably related to marginal job performance. Accordingly, the 2,749 subjects rated in the current study represented "survivors" in terms of job performance, limiting the range of variation likely to be observed in rated work performance data. In this sense, the current findings are conservative to the extent that they probably underestimate the magnitude of the true relationship between predicted job satisfaction and work performance.

In conclusion, the findings of the current study are consistent with those of other studies that have examined the relationship between job satisfaction and work performance. These studies, like the present one, typically find a positive, but weak, relationship between job satisfaction and work performance (Cf. Seashore & Taber, 1975).

References

Alley, W.E., & Matthews, M.D. The Vocational Interest-Career Examination: A description of the instrument and possible applications. Journal of Psychology, 1982, 112, 169-193.

Alley, W.E., Wilbourn, J.M., & Berberich, G.L. Relationships between performance on the Vocational Interest-Career Examination and reported job satisfaction. AFHRL-TR-76-89, AD-A040 754. Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1976.

Matthews, M.D. Vocational interests, job satisfaction, and turnover among Air Force enlistees. Proceedings, Fourth International Learning Technology Congress and Exposition. Orlando, Florida, February 1982(a).

Matthews, M.D., & Berry, G.A. A preliminary look at vocational interest and attrition among Air Force enlistees. Proceedings, Eighth Psychology in the DOD Symposium. Colorado Springs, Colorado, April 1982.

Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. Review and conceptual analysis of the employee turnover process. Psychological Bulletin, 1979, 86, 493-522.

Seashore, S.E., & Taber, T.D. Job satisfaction indicators and their correlates. American Behavioral Scientist, 1975, 18, 333-368.

Waters, L. K., & Roach, D. Job attitudes and predictors of termination and absenteeism: Consistency over time and organizational units. Journal of Applied Psychology, 1973, 87(3), 341-342.

Productivity and Consulting: A New Look at Objective Measures

by

Dr. Hubert S. Feild, Auburn University
Capt Janice M. Hightower, LMDC, Maxwell AFB, AL

for the

Leadership and Management Development Center

The LMDC Consultation Process

The principal goal of LMDC is to help make the USAF a more effective fighting force by focusing on the identification and solving of leadership and management problems, particularly, "people problems." ~~(LMDC, 198[illegible], p. [illegible])~~. LMDC addresses this goal by using organizational development leadership and management problems involving key organizational processes. The resolution of these problems is sought through five basic steps (see Mahr, 1982; Westover, 1979; Hendrix and Halverson, 1979; Short and Hamilton, 1981; and Short and Wilkerson, 1981 for more information).

1. Invitation. Request by the client organization for consultation services.

2. Data Collection. During the first visit, the Organizational Assessment Package (OAP) is administered to the client organization.

3. Data Analysis. After returning to LMDC, the OAP data is analyzed.

4. The Tailored Visit. A second visit to perform the actual consultation, the contents of which is determined by the results from the OAP data analysis.

5. Follow-up. A third visit several months later to re-administer the OAP to measure any change produced by the consultation visit.

The most prevalent research design used by LMDC to evaluate its consultation efforts is the one-group pre-test/post-test design (Campbell and Stanley, 1963), in which the OAP administered during the data collection is the pre-test and the OAP administered during the follow-up is the post-test. Unfortunately, this method suffers from several limitations, known as "rival hypotheses," i.e., hypotheses that represent alternative explanations r any organizational change other than the consultation intervention. These include (Campbell and Stanley, 1963, pp. 7-12; and Cook and Campbell, 1979, p. 52)

- History or the simultaneous events occurring during the consultation.

- Maturation or the natural change within an organization that would have normally occurred.

- Testing or "Hawthorne effect" - type reactions.

- Instrumentation. Changes in the survey instrument.

- Regression or the tendency for the extreme values to gravitate toward the mean.

Objectives

The OAP pre-test/post-test comparison has served as the principal basis for assessing the impacts of a consultation intervention. These perceptual measures are certainly important for characterizing changes in the quality of working life of Air Force personnel. However, these soft measures alone do not give a complete account of the consultation effects and suffer from the limitations described above. Performance data are also needed for providing a clearer picture of the effects of the consultation efforts. Before obtaining this performance data, the following questions must be answered.

- What performance measures should be used?

- Where can these data be obtained?

- What type of research design is most appropriate?

Criteria for Selecting Hard Measures

By drawing upon the findings of Tuttle (1981) as well as other investigators addressing the notion of criteria (Hurst, 1980; Joint Financial Management Improvement Program, 1977), the following criteria were developed for selecting hard measures:

1. Reliability. The measures should provide information that is dependable and accurate.

2. Quantifiable. Quantitative measurement data are more desirable than qualitative data.

3. Available on a frequent basis. Measurement data should be available on at least a weekly or monthly basis.

4. Ease of retrieval. Measurement data should be easily retrievable.

5. Compatible with existing information sources. Measurement data should be from existing information sources rather than from new data sources.

6. Sensitive to change. The measurement data must be sensitive to detect and discriminate among differences in performance, yet not so sensitive as to be influenced by external factors.

7. Controllable by client group. Members of the organization under study should be able to affect the outcome being measured.

8. Uniqueness. Multiple measures of organizational performance are needed to adequately capture an organization's effectiveness and efficiency.

9. Comparable. Measurement data should be comparable from one time period to another.

10. Validity. Measures chosen should assess what they are supposed to measure.

An Alternative Design: Interrupted Time Series

Because of the nature of USAF hard measures, i.e. most are reported over time and by work unit (not by individual), and because of the weaknesses of the pre-test/post-test design, a new research design is proposed: interrupted time series. Interrupted time series was selected for three reasons.

1. The nature of USAF hard measures, i.e., most data are collected and reported over time.

2. In evaluating a time series prior to and after an intervention, several types of effects in the series are tested:

   - A change in the level of intercept of the series.

   - Changes in the slopes of a series may be tested.

   - Effects can also be studied with respect to whether they are continuous or discontinuous.

   - Time series effects can be tested in terms of whether they are instantaneous or delayed following an intervention.

3. The only principal threat to internal validity of the interrupted time series design is history.

Application to Consolidated Base Personnel Office (CBPO) and Aircraft Maintenance

Two client groups, the CBPO and Aircraft Maintenance, were chosen to investigate the feasibility of using existing performance-based measures for evaluating consultations. For the CBPO analysis the Proficiency Status Reporting System (P-Status) was selected as a source for performance data. For the Aircraft Maintenance analysis the Maintenance Data Collection (MDC) and the Maintenance Management Information Control System (MMICS) were selected as a source for performance data.

The P-status report seems to be a useful vehicle for providing some hard measures on CBPO performance. Within one week, monthly P-status reports were received from a large, mid-western Air Force base, suggesting that it would be possible to obtain relevant hard measures through the mail. Two measures appeared to be particularly useful, i.e. Late Airman Performance Reports and Late Officer Effectiveness Reports. The utility of two other measures (Personal Reliability Program and Testing No-Shows) may vary across particular Air Force bases.

The quality of maintenance data studied was somewhat disappointing. There was no problem in obtaining the quantity of data necessary in a timely manner (within one week). Preliminary review of nine measures showed that five measures might be suitable. After further analysis, three of the five had unacceptably low reliability estimates, while the remaining two (Partial Mission

Capable--Maintenance and Scheduling Effectiveness) had marginally acceptable reliability estimates (odd-even, Spearman-Brown corrected reliability estimates of .74 and .63, respectively).

Recommendations

The preceding analyses and results lead to several recommendations for LMDC to consider in implementing its evaluation of the consultation efforts using hard measures.

1. Time series design and analysis should be employed as a method for analyzing hard measures.

2. In addition to data availability, the hard measures chosen should meet as many of the criteria for hard measures outlined in this paper as possible.

3. "Tailored" criteria may have to be developed to adequately assess organizational changes.

4. Late Airman Performance Reports and Late Officer Effectiveness Reports could be used to evaluate the effect of the consultation effort on the Quality Force section in CBPO.

5. Partial Mission Capable Rate--Maintenance and Scheduling Effectiveness could be used to evaluate the effect of the consultation effect on an Aircraft Maintenance organization.

6. Future research should investigate the applicability of time series analysis to previous LMDC consultations.

7. LMDC should carefully examine the utility of those measures developed by Air Force productivity researchers for possible inclusion in its evaluation program.

8. Future research should examine the relationships among various hard measures and the soft measures of the OAP.

References

Campbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1979.

Cook, T. D. and Campbell, D. T. Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally, 1963.

Hendrix, W. H. and Halverson, V. B. Organizational survey assessment package for Air Force organizations (AFHRL-TR-78-93). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, 1979.

Hurst, E. G. Attributes of performance measures. Public Productivity Review, 1980, 5, 5-17.

Joint Financial Management Improvement Program. Implementing a productivity program: Points to consider. Washington, D. C.: Author, 1977.

LMDC. History and structure of the Organizational Assessment Package survey. Maxwell Air Force Base, AL: Leadership and Management Development Center, 1981.

Mahr, T. A. Manual for the Organizational Assessment Package survey. (Technical Report 82-1560). Maxwell Air Force Base, AL: Leadership and Management Development Center, 1982.

Short, L. O. and Hamilton, K. L. An examination of the reliability of the Organizational Assessment Package (OAP). Maxwell Air Force Base, AL: Leadership and Management Development Center, 1981.

Short, L. O. and Wilkerson, D. A. An examination of the group differences aspect of construct validity of the Organizational Assessment Package. Maxwell Air Force Base, AL: Leadership and Management Development Center, 1981.

Tuttle, T. C. Productivity measurement methods: Classification, critique, and implications for the Air Force (AFHRL-TR-81-9). Brooks Air Force Base, TX: Manpower and Personnel Division, 1981.

Westover, T. O. Consultant's guide to the management consultation process. Maxwell Air Force Base, AL: Air Command and Staff College, 1979.

AD P000811

The Measurement of Organizational Productivity
A Description and Field Test

Michael D. Matthews
Charles N. Weaver

Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235

Thomas C. Tuttle
Maryland Center for Productivity
and Quality of Working Life
University of Maryland
College Park, Maryland 20742

Among the most serious obstacles to the study of productivity is the so-called criterion problem, that is, the measurement or assessment of productivity itself. Most published studies of productivity rely on indirect methods of productivity measurement, mainly on the perceptions of productivity reported by supervisors and job incumbents (eg., Berry and Matthews, 1982; Field and Hightower, 1982). Few investigations employ productivity criteria which have resulted from engineering studies (cf. Tuttle, 1981). As a consequence of the widespread use of subjective criterion measurement, the results of many studies of productivity are of questionable validity and limited generality.

Like many civilian organizations, the Air Force is concerned with enhancing and monitoring organizational productivity. And like productivity research in the civilian sector, attempts by the Air Force to measure productivity have been hampered by the lack of objective criterion measures. The lack of objective measures, especially in organizations where engineering-based criteria are not possible, led the Air Force to sponsor the development of a procedure for generating objective measures of productivity. This procedure, referred to as the Methodology for Generating Effectiveness and Efficiency Measures (MGEEM), was developed and described by Tuttle (1981). The inclusion of the words "efficiency" and "effectiveness" in the name of the methodology refers to the notion proposed by Tuttle (1981) that productivity involves considerations of both of these components. That is, productivity is defined as the volume of resources used to provide products and services (efficiency) and the extent to which these products and services conform to acceptable standards of mission performance (effectiveness).

The current paper summarizes the results of a field test of the MGEEM methodology. The field test, conducted in three different Air Force organizations, was undertaken to determine (1) the extent to which the MGEEM and its products are acceptable to organizational participants, (2) the generality across similar organizations of productivity indexes developed using the methodology, and (3) the extent to which indexes developed are cost effective as indicated by their use of existing data. A more complete discussion of the design, results, and implications of this study is available in Tuttle, Wilkinson, and Matthews (in press).

Method

Target Organizations. Three Air Force functions with different missions were studied: (1) Weather, (2) Administration, and (3) Propulsion. Eight organizations in each of the three functions were drawn from the following Air Force commands: SAC, TAC, MAC, ATC and ARC. In all, 24 organizations and 11 bases were included in the field test.

MGEEM Methodology. The MGEEM involves a group decision making process known as the Nominal Group Technique (Delbecq, Van de Ven, and Gustafson, 1975). The Nominal Group Technique (NGT) consists of six steps: (1) silent generation; (2) round-robin listing of ideas generated by individual group members; (3) discussion and clarification of the raw list of ideas developed; (4) individual voting to prioritize items from the list; (5) further voting and clarification of items and voting patterns; and (6) additional voting and discussion, if necessary to achieve consensus. The NGT requires the use of a skilled group facilitator to conduct the group process. While the facilitator guides the group in making a decision, he/she must not attempt to lead the group toward any particular decision.

Procedure. The NGT process, as utilized in this field test, was used to generate indexes of organizational productivity. Two types of indexes were generated: (1) Key Results Areas (KRAs); and (2) Indicators. In order to generate these indexes, two groups of organizational members were involved in the NGT process. The first group, Group A, consisted of the organization's commander and representatives from the next lower level of management. Group A was tasked with the development of KRAs for the organization. KRAs were generated in response to the question "What does the Air Force pay this organization to do?" KRAs were proposed by members of Group A on the basis of their belief that the KRA tapped the basic mission-essential goals or products of the organization. Group A was then asked to vote and prioritize in order to develop a list of six to nine KRAs, the number depending on the diversity of the organization's mission and on the time available to conduct the NGT process.

Following the generation of KRAs by Group A, Group B was formed. Group B consisted of all members of Group A (except the commander) and their immediate subordinates. Group B was tasked to develop six to nine "Indicators" of efficiency and effectiveness for each KRA.

Each organization in the study was visited by a single researcher (facilitator) for five days. On the first day, an inbriefing and familiarization with the subject organization was conducted. On the second day, Group A was formed and KRAs were developed. Days three and four involved the generation of Indicators by Group B. Day five consisted of a review of the KRAs and Indicators with the commander of the organization. In addition, this discussion with the commander identified existing data sources which could provide information required to form the Indicators in actual operational use.

Results

Three questions were addressed in the field test of the MGEEM. First, to what extent did the MGEEM generate indexes which were acceptable to personnel in the organizations studied? Second, how consistent were the generated indexes among organizations within functions? Third, to what extent did the generated indexes make use of existing data?

Acceptability of Indexes

In order to determine the acceptability of the results of the MGEEM procedure to organizational participants, a Participant Feedback Report (PFR) was completed by each participant following their experience in the MGEEM process. The PFR is described in detail by Tuttle et al. (in press). Analysis of responses to the PFR by Group A showed a clear understanding of the purpose of the process and a favorable reaction to the facilitator, as well as to the working climate created. Furthermore, the consistency of these three findings across functional areas of Weather, Propulsion, and Administration provided an affirmative answer to the question of whether the facilitators conducted the process similarly in the three functions.

Group A participants viewed their task as only moderately difficult, but interesting. The KRA indexes were considered acceptable to Group A members as was the priority ranking of KRAs. Group A rated itself as very successful and rated the success of the total MGEEM process as only slightly less than very successful. For all three functions (Weather, Administration, and Propulsion), the members of Group A reported an increase in productivity awareness as a result of participating in the process.

The results for Group B were very similar to the results of Group A. Group B members expressed satisfaction with their success in generating indicators and felt that the process was beneficial in helping them understand their organization's mission. As with Group A, Group B members expressed satisfaction with role of the facilitator, the work climate created, and with the process used. They, too, found their task only moderately difficult, but interesting. Compared to Group A, members of Group B expressed a slightly lower initial level of productivity awareness, but similarly felt that their participation in the process raised their level of awareness concerning productivity.

With only one exception, the MGEEM process was viewed favorably by unit commanders. Other management and non-management participants, as a group, felt that the process and its results were quite acceptable. Thus, in terms of participant reactions, the MGEEM process was generally viewed as quite acceptable.

Similarity of Indexes Within Functions

The similarity of indexes between pairs of organizations within the same function, that is, between two Weather detachments, or two Administration or Propulsion organizations, was investigated in two ways. First, each unit commander and his/her deputy were asked to make judgments about differences between KRA's and indicators for their organization in comparison with KRAs

and Indicators for each of the other organizations in their functional area. Second, the researchers rated the similarity between all possible pairs of organizations within similar functions in which they acted as facilitator of the MGEEM process. Both groups of raters, the unit commander and his/her deputy, and the researchers identified KRA's and Indicators which were the same or substantially the same in each pair of organizations considered. Examples provided in the instruction booklet were designed to define "same" or "substantially the same" at the same level of generality, the same meaning, and the same item form (e.g., ratio, error count, etc.).

Prior to the series of similarity analyses conducted by commander/deputies and researchers, it was hypothesized that the three functional areas, Weather, Administration, and Propulsion, would differ in terms of average pair-wise similarity of Indicators. Influences which were hypothesized to contribute to these differences included command differences, differences in the extent to which performance measurement is institutionalized within the function, homogeneity of the organizations, and the differences produced by the facilitators.

Considering the influences hypothesized to contribute to organizational differences, it was predicted that the Weather function would be the most homogeneous of the three functional areas. In contrast to the Administration and Propulsion functions, all Weather organizations belong to a single command, measure many facets of their performance as a common practice, and (although personnel in Weather organizations can fall into three different job types) do work which is highly interrelated and has a common focus. The next most similar Indicators were predicted to be in the Propulsion function. While Propulsion organizations cut across three commands, the work is quite similar, performance measurement is used extensively, and the work performed is perhaps the most homogeneous of the three functions studied. The lowest similarity Indicators were predicted for the Administration function which spans three commands and does not measure performance to the extent of the other two functions. In addition, their work is separated into three very distinct job types which are always geographically separated. Finally, two facilitators were employed in the work with Administration organizations while only one was employed in Weather and Propulsion.

Results of the similarity analysis showed differences among the three organizations in average similarity ratings by both participants and researchers, but the differences were in the hypothesized direction. Average similarities for KRAs for Administration, Propulsion, and Weather were, respectively, 37.8, 58.9, and 48.6 percent for participants and 21.6, 35.1, and 46.5 percent for researchers. Average similarities for Indicators for the three organizations were, respectively, 10.8, 18.8, and 18.8 percent for participants and 6.1, 11.9, and 18.9 percent for researchers. These results support the hypothesized predictions that the ranking of similarity ratings would be in the following order: (1) Weather, (2) Propulsion, and (3) Administrative. One exception, the case of participant ratings of Propulsion, can be discounted because the ratings are based on a very small subset of the sample.

Cost Effectiveness of the Indicators

A third important aspect of the field test evaluation concerned the extent to which Indicators generated in the MGEEM process can be formed using existing data. In the organizations studied, there are at least three forms of existing data. The most obvious is an entry on an existing reporting

form. Another type of existing data includes entries in management information system products provided to managers/commanders by staff support agencies or higher headquarters. Finally, there is a variety of local data, such as status boards, customer feedback forms received, and duty rosters. In the latter case, data are available but may not be tabulated in the exact format required to form the Indicator. Nevertheless, all three categories are grouped together for this analysis under the heading "existing data." If an Indicator requires that a new log be established or that some other form of initial data collection be instituted, then the Indicator is considered not to make use of existing data. Inclusion of Indicators or KRAs on the final list does not necessarily mean that data presently exist for their support. It may be that the unit commander has determined that the Indicators or KRAs are sufficiently important to justify the cost of collecting the additional data.

Results of the analysis of Indicators with respect to use of existing data showed that for both Administration and Propulsion the percentages of Indicators that require no new data collection exceeded 90% and that for Weather the percentage was 80 percent. Thus, from the viewpoint of the cost-effectiveness of the Indicators, the Indicators generated may be said to have made extensive use of existing data.

Discussion

A field test of the MGEEM methodology demonstrated that (1) the process was highly acceptable to participants, (2) the judged similarity of KRAs and Indicators varied from low to moderate across organizations within the three functions, and (3) the Indicators developed were cost effective. These findings have implications for both research applications of the methodology and for organizational productivity measurement and enhancement applications.

With reference to implications for research applications of the MGEEM methodology, the limited generality of indexes across organizations within the same function would tend to restrict the value of the methodology to the extent that productivity indexes relevant to one organization would not apply to similar organizations. However, as discussed more thoroughly by Tuttle et al. (in press), evidence suggests that two refinements of the methodology may result in a level of similarity across organizations which will be acceptable for research purposes. The first refinement would be to allow more time in KRA and Indicator development. The second refinement would be the addition of another step in the MGEEM procedure in which idiosyncratic and unit-specific indexes would be eliminated before indexes are compared across organizations. Given these refinements of the procedure, and in view of the fact that most Indicators generated utilize existing data sources, the MGEEM methodology would seem to hold promise as a research tool for measuring productivity across organizations.

The results of the field test clearly demonstrate that the MGEEM methodology is useful in generating productivity indexes for uses within organizations. These uses, both diagnostic and therapeutic in nature, do not appear to be affected by limitations in inter-organizational generality. The high acceptability of the methodology to field test participants and its apparent ability to utilize existing data sources for most Indicators developed underscore the potential utility of the methodology as a management tool.

In conclusion, the field test of the methodology demonstrated that the MGEEM can be applied in operational Air Force units, is very acceptable to participants, and generates useful productivity indexes that are readily obtainable from existing data. The methodology generates organizational productivity indexes useful for both productivity enhancement and monitoring. And, assuming that proposed refinements are incorporated into the process, the MGEEM should generate productivity indexes useful in inter-organizational applications.

References

Berry, G. A., and Matthews, M. D. Vocational interests and job performance among first-term Air Force members. Proceedings, 24th Annual Conference of the Military Testing Association. San Antonio, November 1982.

Delbecq, A. L., Van de Ven, A. H., and Gustafson, D. H. Group Tecnniques For Program Planning. Glenview, IL.: Scott, Foresman and Company, 1975.

Field, H. S., and Hightower, J. Productivity and consulting: A new look at objective measures. Proceedings, 24th Annual Conference of the Military Testing Association. San Antonio, November, 1982.

Tuttle, T. C. Productivity measurement methods: Classification, critique, and implications for the Air Force. AFHRL-TR-81-9, AD-105267. Brooks Air Force Base, Texas: Manpower and Personnel Division, Air Force Human Resources Laboratory, September, 1981.

Tuttle, T. C., Wilkinson, R. E., and Matthews, M. D. Field test of a methodology for generating organizational productivity indicators. AFHRL-TR. Brooks Air Force Base, Texas: Manpower and Personnel Division, Air Force Human Resources Laboratory, in press.

Effects of Feedback and Goal Setting on Productivity

by

Robert D. Pritchard
Department of Psychology
University of Houston

The research described here is the most recent effort in a program of research sponsored by the Air Force Human Resources Laboratory and the Air Force Office of Scientific Research. The basic logic of this program is that it is appropriate to explore ways of increasing productivity which can be implemented by local management and which rely on intrinsic motivation to increase productivity.

In the first phase of this program of research, the existing literature was examined to isolate those variables that had promise for affecting intrinsic motivation (Pritchard & Montagno, 1978).

In the second phase, some of these variables were explored in a controlled setting to begin to assess their suitability for eventual field application. Feelings of personal control and competence, as well as contingent extrinsic rewards, were examined by Fisher and Pritchard (1978). Performance feedback was addressed by Pritchard and Montagno (1978).

The third phase attempted to isolate variables which could be implemented in an operational Air Force environment and to test a fairly large number of different possible applications in a controlled, yet realistic setting. In this stage, it was necessary to narrow the list of potential determinants of intrinsic motivation to a smaller subset for more careful study. After evaluating them in terms of (a) their potential use in a field setting, (b) the feasibility of testing them in the work simulation setting to be used, and (c) the quality and quantity of previous literature available, major emphasis was placed on the performance feedback variable. Six dimensions of feedback and a job design variable, completeness of the task unit, were evaluated in the controlled setting. The major conclusion of this study was that feedback had meaningful potential for increasing productivity (Pritchard, Montagno, and Moore, 1978).

In the most recent phase, described here, several specific types of performance feedback, singly and in conjunction with goal setting, were selected to be tested in an operational work environment similar to those found in some Air Force settings. A more complete report on this study may be found in Pritchard, Bigby, Beiting, Coverdale, and Morgan (1981).

Procedures

Two civilian clerical type jobs were selected for study. The experimental conditions consisted of various types of feedback, and one type of goal setting. Based on our previous research, the optimal type of feedback was identified as being 1) individual in nature in that each employee was given

feedback on his/her own performance, 2) private as opposed to public in nature, and 3) directed to the specific tasks performed by the employee. However, there were two other dimensions of feedback which the previous work had shown to be equally effective and, as such, were directly examined in the project. The first was personal vs. impersonal feedback. In personal feedback, the information came from the supervisor, it was clear to the subordinate that the supervisor had seen the feedback, and it was evaluative in nature. That is, there was a good-bad component to the feedback. In impersonal feedback, the information did not come directly from the supervisor, but rather from other sources. In this type of feedback it was not clear that the supervisor had seen the information, and the information was purely descriptive of performance rather than evaluative.

The second feedback dimension was absolute vs. comparative. In absolute feedback the employee received information only about his/her performance. In comparative feedback the employee also received information about how he/she performed compared to the rest of the work group.

In order to implement these feedback procedures, computer software was developed to produce daily feedback reports for each employee. These reports indicated the employee's performance on the various types of tasks for the most recent day that could be processed. (It typically took 2-4 days to process the reports.) In addition, each employee was given his/her average performance scores for the previous week.

The second type of condition was goal setting. To institute goal setting, supervisors were trained to assist their subordinates in setting specific, moderate to difficult goals in a manner that would promote employee acceptance of the goals. A set of easy, moderate, and difficult goals were given to the supervisor for each employee. These suggested goals were personalized for each employee by consideration of his/her place on the learning curve and his/her potential for improvement. Supervisors met with their subordinates at regular intervals to set or reset goals throughout the goal setting condition.

Design

In both jobs there was a day shift and an evening shift. Each shift was treated as a separate experimental group. For each, there was a baseline period, during which performance data were collected but no experimental conditions were administered, followed by a first treatment and a second treatment. All first treatments involved some combination of feedback conditions. In three of the four groups, goal setting was added to feedback in the second treatments. In the fourth, the type of feedback was changed. Such a design allows for a direct comparison of the effects of the feedback and goal setting procedures on productivity.

Results

1. The treatments showed an overall positive effect on performance. Increases in quantity of output typically ranged from 5% to 10% with a mean increase of 6.4%. Error rates decreased. The mean decrease in

errors was 11%, with over half the decreases in the 15% to 28% range.

2. Personal feedback was equally as effective as impersonal feedback.

3. Absolute feedback was equally as effective as comparative feedback.

4. Goal setting plus feedback showed higher performance than feedback alone.

5. The positive effects of the treatments did not diminish over time.

6. The treatments had fairly strong effects on employees who were initially low performers. They did not have much effect on employees who were initially high performers.

7. There was some evidence that the treatments effected the rate of learning, but these results were not present in all situations.

8. Attitudes under the treatments were as good or better than before the treatments.

9. Reactions of the unit supervisors were very favorable. They felt that their subordinates' productivity and attitudes improved and saw the feedback and goal setting procedures as an excellent management information and counseling tool.

Conclusions

It was concluded that feedback and feedback plus goal setting are very useful techniques that could be used in field settings to improve productivity.

References

Fisher, C.D., & Pritchard, R.D. Effects of personal control, extrinsic rewards, and competence on intrinsic motivation. AFHRL-TR-78-20, AD-A058 417. Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, July 1978.

Pritchard, R.D., Bigby, D.G., Beiting, M., Coverdale, S., Morgan, C. Enhancing productivity through feedback and goal setting. AFHRL-TR-81-7, Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, July 1981.

Pritchard, R.D., & Montagno, R.V. The effects of specific vs. non-specific and absolute vs. comparative feedback on performance and satisfaction.

AFHRL-TR-78-12, AD-A055 693. Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, May 1978.

Pritchard, R.D., Montagno, R.V., & Moore, J.R. Enhancing productivity through feedback and job design. AFHRL-TR-AD-A061 703. Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, August 1978.

AD P000813

QUALITY CIRCLES IN THE DEPARTMENT OF DEFENSE: SOME PRELIMINARY FINDINGS

Robert P. Steel, Nestor K. Ovalle, 2d, and Russell F. Lloyd
Air Force Institute of Technology

ABSTRACT

Quality Circles management has been greeted with tremendous enthusiasm by American managers attempting to emulate the recent economic success of the Japanese industrial complex. The Department of Defense is becoming actively involved in Quality Circles. Scant rigorous research exists on the effectiveness of Quality Circles as a management tool. A nonequivalent control group design compared 14 Quality Circles groups to 37 untrained work groups on a number of attitudinal measures. No consistent differences between groups were detected. Design flaws limiting the validity of the findings included sample-size problems, experimental mortality, weak treatment effects, and poor experimental control. The study's results should be treated as highly tentative and further research should attempt to overcome these design limitations.

The popular management literature is replete with testimonials praising Quality Circles management as a revolution in the management of work organizations. Quality Circles are designed to foster work group-oriented decision making geared to the solution of task-related problems. Commonly, face-to-face work groups (usually 5-12 people) will meet periodically to identify problems relating to the productivity of the unit or to the quality of outputs produced. A preprogrammed set of decision-making tools including brainstorming, cause-effect analysis, pareto diagramming, and the like are routinely used as guides to problem analysis (Rehg, 1976).

The tremendous enthusiasm greeting Quality Circles management is indicative of the interest among American managers in Japanese management techniques. Quality Circles management took root in Japan some years ago and is now being offered as a partial explanation for the productivity gains realized by Japanese industry relative to the world economy as a whole.

Beyond the realm of opinion and anecdotal evidence, very little systematic and controlled evaluative research on the effects of Quality Circles programs currently exists. With few exceptions (e.g., Hunt, 1981; Tortorich, Thompson, Orfan, Layfield, Dreyfus, & Kelley, 1981), there has been little published work evaluating the outcomes of Quality Circles interventions in order to ascertain their effect upon attitudinal or behavioral criteria. The present study reports the results of a six-month longitudinal investigation examining attitudinal changes in Quality Circles members as a function of participation in Quality Circles groups.

Exact figures are not available, but estimates of the number of Quality Circles operating in the Department of Defense indicate that as many as 1,000

Quality Circles (Mento, Note 1) may currently exist within the various military departments. Such an investment of resources should be counterbalanced with a serious commitment toward research and evaluation examining the results of Quality Circles activities for groups instituted in the federal sector.

This paper describes a longitudinal evaluation carried out jointly by the Leadership and Management Development Center (LMDC) and the Air Force Institute of Technology on the effects of Quality Circle participation.

METHOD

Subjects

Between the periods 10 Sep 80 and 1 May 81, six Quality Circles were inaugurated in the Civil Engineering Division of a Department of Defense installation. These groups served as the source of data for the present study. Typically, groups are provided with an orientation and some initial training on the merits/techniques of Quality Circles followed by regular meetings designed to identify and resolve work problems. A total of 383 individuals responded during the final wave of survey data. The departments involved in the Quality Circles effort ranged in size from 3-21 assigned employees. The average size of the departments involved was 10 employees.

Measures

The Organizational Assessment Package (OAP) was used to assess attitudinal and cognitive changes in study participants. The OAP is a survey questionnaire containing 109 items measuring employee attitudes (e.g., job satisfaction, organizational climate), beliefs (e.g., work-group productivity, job characteristics), behavioral intentions (career intentions), and demographic characteristics (e.g., sex, pay grade, length of service). Except for the demographic factors which are distributed on both ordinal and nominal scales, all items are arrayed on seven point Likert-type scales. The non-demographic items in the OAP are keyed to 23 underlying psychological factors which were identified through factor analysis. Developmental procedures, factor analytic results, and scale reliabilities for the OAP may be found in Hendrix (1979) and Hendrix & Halverson (1979).

Procedures

The OAP was administered to the entire Civil Engineering organization by LMDC in September 1980 (pretest) and again in May 1981 (posttest). The entire organization was surveyed in order to provide a control group against which the Quality Circles groups might reasonably be compared.

The study design approximates a nonequivalent control group design (Campbell & Stanley, 1963) and is described in more detail elsewhere (Steel, Lloyd, Ovalle, & Hendrix, 1982).

The experimental treatment condition (called the Quality Circles group) contained results (aggregated by department) for each of the fourteen departments active in the Quality Circles program. These data were pooled from the responses of 133 individuals (posttest). Data were not aggregated according to actual Quality Circles boundaries because some circles crossed formal departmental lines. A control condition was composed of the departmental means for 37 work units (250 individual respondents on the posttest) that did not directly participate in the Quality Circles process.

RESULTS

Demographic Measures

Mean difference tests (t-tests) between the Quality Circles groups and the control groups on selected demographic variables are displayed in Table 1.

TABLE 1

Quality Circles and Control Group Demographic Characteristics

	Pretest			Posttest		
	Quality Circles	Control Group		Quality Circles	Control Group	
Variable	$\overline{X}$	$\overline{X}$	t	$\overline{X}$	$\overline{X}$	t
Age	30.42	36.91	2.51*	32.00	37.61	1.80
Pay Grade	4.86	6.2	2.84**	5.27	5.51	.43
Years in Air Force	4.52	4.77	.74	5.03	5.13	.26
Months in Present Field	6.01	6.11	.39	6.04	6.01	.09
Months at Current Station	5.09	5.63	1.72	5.34	5.23	.34
Months in Present Position	4.20	5.01	2.61*	4.57	4.31	.77
Education Level	2.41	3.01	2.50*	2.56	2.76	.91

*$p < .05$
**$p < .01$

Several significant premeasure differences were detected between the treatment and control groups. Control group members appeared to be significantly older, had a higher average pay grade, had performed longer in their current position, and were significantly better educated than their Quality Circles counterparts. Considerable leveling of the sample appears to have taken place prior to the posttest since by this time significant demographic differences between experimental conditions had disappeared.

To further amplify the demographic differences between groups, t-tests were carried out comparing pretest and posttest means within treatment groups. This set of tests was conducted to shed light upon changes in group composition over time. A significant reduction in the average time spent in present position was detected within the control group's pre-post scores ($t=2.61$; $p < .02$). No other significant changes were found.

Attitudinal Measures

Pretest and posttest means for the Quality Circles and control groups are presented in Table 2. To avoid restrictive assumptions associated with analysis of covariance (e.g., homogeneity of regression slopes), the data were analyzed using stepwise hierarchical regression analysis. Posttest scores on the 23 OAP factors were employed as criteria. Pretest results were entered on the first step of the regression analysis to eliminate criterion variance attributable to pretest differences. A dummy variable representing treatment condition (Quality Circles or Control) was entered in step 2 of the analysis. Significant increases in $\underline{R}^2$ on step 2 would indicate explanation of unique criterion variance attributable to the Quality Circles intervention. No significant increases in $\underline{R}^2$ were observed for the regression on the 23 OAP attitudinal measures. Actual increases in $\underline{R}^2$ observed when treatment condition was entered into the regression equations ranged between .000 and .046.

These results tend to suggest that participation in the Quality Circles program at this installation had minimal impact on the attitudinal responses of participants during the period of study. This conclusion must be regarded as highly tentative, however, because several technical limitations operated to severely confound study results.

TABLE 2

QUALITY CIRCLES AND CONTROL GROUP MEANS FOR OAP FACTORS

	Pretest		Posttest	
FACTOR	Quality Circles $\bar{X}$	Control Group $\bar{X}$	Quality Circles $\bar{X}$	Control Group $\bar{X}$
Skill Variety	4.63	4.75	4.75	4.89
Task Identity	4.71	4.96	4.96	5.14
Task Significance	5.50	5.39	5.41	5.66
Job Feedback	4.65	4.79	4.71	4.79
Work Support	4.11	4.14	4.40	4.39
Need for Enrichment Index	5.25	5.53	5.20	5.36
Job Performance Index	4.54	4.58	4.72	4.65
Pride	4.79	4.78	4.96	5.08
Task Characteristics	4.88	4.99	5.00	5.11
Task Autonomy	3.9[illegible]	4.41	4.12	4.45
Work Repetitiveness	5.03	4.82	5.07	4.92
Desire for Repetitive Tasks	3.79	3.23	3.46	3.27
Advancement/Recognition	3.40	3.80	4.07	4.06
Supervision	4.58	4.44	4.86	4.98
Supervisory Communication Climate	4.30	4.18	4.56	4.44
Organizational Communication Climate	4.18	4.56	4.23	4.62
Work Group Effectiveness	5.07	5.43	5.34	5.35
Job Satisfaction	4.81	5.09	5.11	5.29
Job Training	4.19	4.59	4.55	4.61
General Organizational Climate	4.45	4.76	4.47	4.37
Job Motivation Index	98.26	113.98	113.38	119.71
OJI Total Score	63.53	66.26	65.64	68.14
Job Motivation Index (Additive)	13.37	14.18	13.91	14.45

DISCUSSION

Quality Circles management enjoys immense popularity and interest at the present time. Conventional wisdom holds that this technique can be a very effective means of enhancing work group effectiveness. Carefully conducted scientific research is sorely needed to evaluate the effects of Quality Circles participation upon the attitudes and behavior of their members. The present study attempted to make a small contribution toward filling that void.

Taken as a whole, the configuration of results tend to support the conclusion that the Quality Circles groups initiated in this organization had little, if any, influence upon the constellation of work related attitude measures contained in the OAP. Conclusions from the present study's findings must be tempered considerably by the recognition of a number of technical and design limitations which may have served to diminish the validity and generalizability of the study's results.

These methodological difficulties are enumerated as they represent significant obstacles which future Department of Defense research on Quality Circles must attempt to overcome in order for meaningful unambiguous evaluation to be possible.

Five methodological impairments confounded study results. (1) Some Quality Circles studied did not have an opportunity to reach full maturity prior to collection of the postmeasure. Quality Circles groups at this installation did not all begin at the same time. Rather, start-up dates for the various groups were staggered throughout the observation period. In fact, three of the Circles in this study had less than a month to develop prior to administration of the posttest. (2) Experimental mortality altered the character of samples in both treatment conditions. Significant fluctuations in the demographic measures over time indicate that there may have been changes in the composition of treatment groups during the course of study. This could occur through such mechanisms as employee turnover, new hirings, transfers, or reassignments. Incomplete exposure to the Quality Circles treatment for some experimental subjects would tend to water down treatment effects and lead to a lack of significant group differences. (3) The treatment groups were not equivalent at the outset of the study. Significant differences between the treatment groups on the demographic measures at the pretest were observed. Statistical control (controlling for pretest differences) is a less than perfect control for pre-existing differences between groups in a study as uncontrolled differences may interact with the treatment to produce uninterpretable findings. (4) Nonattitudinal measures of outcomes were not investigated. Improvements in employee morale have been mentioned as outcomes anticipated from participation in Quality Circles (Dewar, 1980), but behavioral and results criteria should also be examined. (5) The sample size used in this study was small by most standards. The power of statistical tests to detect treatment effects was attenuated by small sample sizes in both treatment conditions and, therefore, some incidence of Type II errors is to be expected.

Future Research

Based upon the many design flaws encountered by this study, the results of this investigation must be viewed as most inconclusive. These design flaws should not be seen as insurmountable deterrents to worthwhile research on Quality Circles. Rather, more carefully controlled research can (and will) be done studying Quality Circles in the Department of Defense. We have expanded our research efforts on Quality Circles and now have evaluations on-going at six different sites. We sincerely hope that present and future research efforts may benefit from some of our "lessons learned."

REFERENCE NOTES

1. Mento, A. Program Director, AFIT Quality Circles Program. Personal communication, August 31, 1982.

REFERENCES

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Dewar, D. L. The Quality Circle guide to participation management. Englewood Cliffs NJ: Prentice-Hall, 1980.

Hendrix, W. H. Organizational assessment indices of effectiveness, (AFHRL-TR-76-46) Brooks AFB TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, December 1979 (NTIS No. AD-A-68-476).

Hendrix, W. H., & Halverson, V. B. Organizational Survey Assessment Package for Air Force organizations, (AFHRL-TR-78-93) Brooks AFB TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, February 1979 (NTIS No. AD-A068-476).

Hunt, B. Measuring results in a Quality Circles pilot test. The Quality Circles Journal, 1981, 4, 26-29.

Rehg, V. What are the tools of the QC circle? In D. M. Amsden and R. T. Amsden (Eds.) QC Circles Applications, Tools, and Theory, Milwaukee WI: American Society for Quality Control, 1976.

Steel, R. P., Lloyd, R. F., Ovalle, N. K., & Hendrix, W. H. Designing Quality Circles research. The Quality Circles Journal, 1982, 5, 40-43.

Tortorich, R., Thompson, P., Orfan, C., Layfield, D., Dreyfus, C., & Kelly, M. Measuring organizational impact of Quality Circles. The Quality Circles Journal, 1981, 4, 24-34.

TABE AND THE COMBAT SOLDIER

Chair: J. E. Gerber

Use of the Test of Adult Basic Educations (TABE) was discussed. Discussion included its application as a placement test in Army Basic Skills Education, and as a counseling aid in developing career education plans. Some findings on correlation studies of scores of soldiers tested with both the TABE and the Armed Services Vocational Aptitude Battery (ASVAB) were presented.

TEST OF ADULT BASIC EDUCATION (TABE OR NOT TABE)

Michael Bachem, Ph.D., Temple University

This paper is being written after thirty-one months working as a representative of an educational contractor, with the responsibility for the Basic Skills Program (BSEP) in the Army's VII Corps. Among the many things I learned was that the contractor does not meddle in the Army's determination of eligibility for BSEP and therefore not in the choice of screening instrument the Army wishes to use. To be sure, the relative merits of various tests were discussed frequently by everyone involved in the administration and the teaching of the BSEP program. What counted for the teachers in our classrooms were our own criterion-referenced diagnostic tests. These tests determined exactly which skills a soldier would need to work on during the sixty hours of instruction. The TABE score with which the soldier entered the class tended to be used as a general indicator of achievement, and was certainly taken into consideration , but the TABE simply does not yield detailed diagnostic information which could then be used to plan a series of learning prescriptions.

During the sixty hours of class, instructors and students concentrated on the skills that had been identified by the diagnostic test, and day to day success, encouragement, and motivation was determined by the degree of mastery of the skills identified on the student's Individual Training Plan. Nevertheless, teachers and students knew that the TABE score achieved after the class would be used as a measure of their success, all of our disclaimers notwithstanding. Surely many students must have wondered why, to oversimplify somewhat, they spent sixty hours doing one thing and then were tested on another. The contractor's representatives also worried much about the dichotomy between their contractual obligation to achieve a 9.0 TABE score and their judgment as professional educators concerning what went into a basic skills program.

This dilemma, which was discussed often, but which neither side seemed to be able to do anything about, is expressed very succinctly by S. Allen Cohen as quoted in The Seventh Mental Measurement Yearbook, edited by O.K. Buro: "The (TABE) battery could be used as a pre-post measurement for groups, but not for individuals." Therefore, one might conclude, what the contractor was doing—not using TABE information for individual diagnosis and prescription—and what the Army was doing—using TABE as a pre-post measurement for large groups—was justified by some of the professional literature. So, perhaps all is well with the way things were done.

There are, however, a few fundamental questions that are raised by the use of TABE in the screening of soldiers for BSEP eligibility. First of all, does the TABE test skills that the Army is interested in? Without presuming to know precisely what the Army, or even the individual commander might be interested in, it is safe to assume that the sills should be those that matter in the performance of a soldier's duty. Without detailed examination of all items on the TABF it is clear that this test must fall short. After all, TABE

is derived from the California Achievement Test Battery, revised to eliminate childish references. But, to quote Cohen again, "What kinds of behavior should a test of literacy for disadvantaged, semiliterate, and illiterate adults tap? Should it assess the same things as are measured in middle-class elementary school children? It is doubtful." Put this way, there is probably wide agreement that a revamped test, such as the TABE, is inappropriate for measuring basic skills needed by active duty soldiers. That the TABE, in its present form, has not been normed, and that the norms are said to be "inherited" from the California Achievement Test, is a dubious, if not unethical practice. The use of elementary school grade levels to categorize adult combat soldiers seems little short of an insult, no matter how desparate their need for remedial work may be.

If one examines the three BSEP curricula, mathematics, reading, and writing, to see how the items on the TABE match the diagnostic tests used by Temple University, one notes clear differences and also some interesting facts. There is most similarity in TABE items and the reading curriculum, less in mathematics, and least in writing. At the same time, TABE gains achieved by a sample of 1518 soldiers over a six months period (July to December 1981) also show some differences, with highest gains in mathematics (1.78), fewer in writing (1.73), and least in reading (1.09), all achieved after sixty hours of BSEP instruction.

The difference is surely not cause by a difference in quality of either the instruction or the curriculum. The differences rather highlight certain simple truths. The subskills in reading are more or less universal. To comprehend language one has to understand sequences of words in context, one has to be able to understand what one reads. This is neither simple, nor easy, but it is rather straightforward, and there is comparatively little opportunity to teach for the test. Therefore, the TABE scores achieved in reading are probably the most reliable, and our reading coordinator expressed no dissatisfaction with the TABE. The TABE reading scores are also the most realistic gains achieved by the contractor in BSEP. The fact that they are also the lowest simply underlines the need for patience in building basic skills in reading.

The TABE mathematics scores are more problematic. The TABE tests skills in computation, in understanding and use of mathematics concepts, and in the ability to work word problems. Some areas, however, are not tested, such as estimation, judging reasonableness of results, and the reasonable use of measurements. Other areas do not receive enough attention, such as understanding and working with percents. These areas were deemed important in our curriculum, and I would presume that they would be important to the job skills of soldiers. Therefore, the TABE results five a somewhat distorted picture of relevant achievements in mathematics. The general applicability of TABE test items to the skills of soldiering remains as problematic with the mathematics section of the TABE as with other curricular areas. This is especially true with the word problems in mathematics, which tend to lack any connection to the work of a soldier.

The TABE mathematics gains were higher than those in any other curriculum. There are surely many reasons for this, some having to do with the ease of reactivating skills learned a long time ago. Other reasons may touch on the

nature of drilling in mathematics, which, if done right, should come more naturally to soldiers than drilling communications skills. It is relatively easy to accept the need to drill computation skills because the common perception is that everyone needs to improve these skills. In contrast, even the lowest achievers on the TABE language test communicate skillfully and with complete mastery among their peers, and therefore tend not to accept the need fro drilling communications skills quite as easily. Another reason may have to do with the notion of accuracy and reliability. It is commonly accepted that the results of a mathematical operation are what they are, and a mathematical proof usually suffices to quiet the skeptics. It is a different matter to convince a reluctant student that generally accepted usage of a certain word or phrase is what it is. Teachers sometimes have to resort to authority without benefit of mathematical proof. Teaching communication classes may approximate the arbitrariness of a foreign language class where rational explanations frequently are simply not available.

The greatest disparity between what the TABE tests and what the contractor taught exists in the area of writing. In a nutshell, the problem exists because you can only test writing skills by evaluating writing samples. The TABE, on the other hand, simply tests communication skills that are easily scorable, primarily capitalization, punctuation, and spelling. Of the 132 communications items on the D level TABE, 104 test these three skills. To be sure, lack of mastery of any of these skills is a powerful stigma in our society, and yet these are essentially editing skills which, I would respectfully submit, all of the participants in this conference are still refining. The more important skills of organization, or sequencing, of separating relevant from irrelevant items, of development, in short of the major ingredients of clear and forceful writing, are not tested by the TABE.

Yet in no other curriculum would it have been as easy to achieve spectacular gains as in the communications area, and in no other area would the soldiers have been as ill served by such a strategy. The spelling lists and the punctuation problems tested on the TABE were available without too much difficulty for many teachers, since the TABE is a commercial item, and instances of astonishing gains would prompt our communications coordinator to suspect that teachers were drilling their students too specifically for the post-test. However easy and tempting this solution may seem, I would assume, and we did assume, that this is not what the Army wanted for its BSEP writing program.

It seems like a truism, but one that has to be recalled occasionally, that one learns to do what one does. What seems obvious in other curriculum areas—one learns comprehension skills by reading, computation skills by computing—is frequently approached too indirectly in the BSEP writing program. It should be clear that, as in all other skill learning, one learns to write by writing and not by practicing editing skills, as the test items on the TABE seem to imply. There is no shortcut to teaching writing by making students write and re-write, daily, if possible, no matter how difficult this may seem. Evaluating actual writing samples produced by students may seem unreliable and subject to arbitrary judgments. But this is precisely where the teacher training efforts of a good contractor would come into play. This training would insist that reliable and fair evaluation of students' writing is possible, and is, in fact, the goal of the writing program.

The reason for this stubborn insistence is, in the final analysis, found in the belief that there is a connection between writing and thinking, and that the better writers will be the clearer thinkers, and that clearer thinking can be taught. In combat there will hardly be time for clear writing, but clear thinking may be a matter of survival.

There are other areas where the TABE falls short of being an adequate screening device. As already mentioned, the use of grade equivalents to express TABE scores makes no sense for this particular population. The test results ought to be stated in value-neutral terms, and, at best, ought to be criterion-referenced. In the best of all possible worlds, the test would be keyed to military occupational skills. Such a test would then also provide initial guidance for teachers even before more specific diagnostic tests can be administered.

Identification of skills needed by soldiers in their MOS would not only provide the basis for better testing, but would also provide the basis for truly job-related curriculum development. Using all the professional help available, its own resources, and its most intimate knowledge of its own requirements, the Army should, in my opinion, develop its own screening device to test soldiers in skills that matter to them, and it should and could, at the same time, define these skills with sufficient clarity so that a truly job-related curriculum could be developed. Failing this, the alternative seems to me to be a BSEP program that removes the requirement of an MOS-related curriculum. In the long run, if some sort of basic skills program should be needed to support combat readiness, an adequate screening device, and an adequate curriculum can only be developed on the basis of thorough cooperation between the Army and a first-rate educational contractor.

(Acknowledgment: This paper would not have been written without the help of friends and colleagues in the Temple University Basic Skills/ESL Program. Special thanks go to Kenneth Schaefer (Communications Coordinator), Jane Paalborg (Mathematics Coordinator), Howard Blake (Reading Coordinator & Curriculum Director), and Frederic Harwood (Curriculum Coordinator).)

(The opinions expressed are the author's own and do not necessarily reflect the opinions of Temple University.)

AD P000815

Mary F. Koss, USAREUR ACES

COUNSELING SOLDIERS ON TABE RESULTS AND CAREER EDUCATION PLANS

INTRODUCTION

The Army's Basic Skills Education Program (BSEP) is designed to "reduce educational deficiencies that hinder soldiers' military duty performance." (AR 621-5, Chapter 2). The key factors in accomplishing this mission are counseling and instruction. This paper will address the counseling function and the use of the Test for Adult Basic Education (TABE) at various stages in that counseling process.

Each Army Education Center is staffed by professional guidance counselors. These counselors accomplish the identification, placement and evaluation of the BSEP process.

There are various methods for identification of soldiers needing BSEP. One of the most common is to review education records for the score achieved on the General Technical (GT) area of the Armed Services Vocational Aptitude Battery (ASVAB). The general rule is that a soldier with a GT score of 90 or below is a potential BSEP candidate.

The second identification factor is a result of the Skill Qualification Test (SQT) taken by soldiers to validate their military occupational specialty (MOS). Those soldiers who do not achieve qualifying scores, particularly in those areas related to reading, writing or mathematics, are referred to the Education Center for counseling and evaluation.

Soldiers who are not functioning adequately on the job may be referred to the Education Center for testing and evaluation by their commanders. These individuals may be enrolled in BSEP classes even though they have GT scores of 90 or above. Commanders may also refer soldiers who must retake the ASVAB in order to reenlist.

Finally, at some installations soldiers are administered the TABE during the in-processing procedure when they arrive at a new installation for permanent assignment.

In each of the first three cases, the soldier is scheduled for the TABE. Either level M or D is used depending on the installation or major command practices/policies. Level M (3-9th grade level) requires a little more time to administer but is more discriminating for those individuals with lower basic skills. Level D (5th to 12th grade level) provides an evaluation measure for soldiers at the upper achievement range and for whom the counselor needs a predictor of success for retesting on the ASVAB or other higher level test/academic experience.

The use of standardized ability tests has come under fire in recent years and some criticisms are undoubtedly deserved. However, the test instruments we have available are only part of a variety of diagnostic tools to provide data on which to base judgements needed to accomplish a mission. The history

of ability testing goes back to the last half of the nineteenth century with the developing of social sciences and probability statistics. World War I gave a major impetus to the field when the need to determine potential success in training for large numbers of young men conscripted into the Army led to the development of the Army Alpha test. David A. Goslin in The Search for Ability (published in 1963 by the Russell Sage Foundation), credited the development and application of the Army Alpha test with setting the stage for "group testing in education...where large numbers of individuals have to be classified quickly and efficiently." (Page 28)

The need for a diagnostic and predictive instrument for a program such as BSEP is obvious. Again, it is only part of the process, one tool among several. The counseling process for BSEP enrollment and subsequent evaluation depends on the professional skill and understanding of the guidance counselor to successfully initiate and monitor BSEP enrollment and progress for the individual soldier.

THE COUNSELING PROCESS

The counseling process begins the same way regardless of the method of initiation. A review of the educational record of the individual and an interview is the the first step. Next the soldier is scheduled for the TABE. The TABE Locator, a short pretest in reading and mathematics skills may be administered first or Level M or D may be scheduled immediately depending on the counselor's evaluation.

Results of the TABE, the soldier's MOS, his or her educational record, the needs of the unit and the availability of instructors are among the factors determining placement in a class. The maximum progress in the time alloted is the aim of the program.

Actual implementation of BSEP may vary; however, optimum time devoted to one subject is four hours per day. Usually classes are taught in three weeks, five days per week for a total of sixty hours. In US Army Europe (USAREUR) this pattern prevails. Classes are taught by credentialed instructors administered by Temple University under contract to USAREUR.

During the course of each class, counselors review the Individual Training Plan (ITP) prepared by the instructors for each student. This review takes place when fifteen hours of instruction have been completed. Progress and the ITP are again reviewed when the class is nearing completion. These reviews and the cooperation between counselor and instructor not only assure that program objectives are being met, but provide incentives to the soldiers. The counseling interview near class completion also allows the counselor to discuss and tentatively estimate the number of additional classes (if any) that the soldier may need to complete the program. The student is considered to have completed the program when they achieve scores above the ninth grade level in all areas of the TABE.

Depending on the student's progress during the class, and the recommenda-

tion of the instructor the student may be scheduled for a post class test with another form of the same level TABE in the subject area covered by the instruction received.

Army Education Center guidance counselors are generally very much aware of the need to provide a supportive environment to the individual soldier when dealing with the sensitive area of academic deficiencies. The use of the term "grade level" must be qualified when working with adults, particularly those who have completed high school, or -- in some cases -- college work. The term "functional grade level" can be utilized to remove some of the negative connotations of a low grade level interpretation from the TABE. In other words, the counselor stresses TABE results as a measuring or comparison score for current skills rather than a school grade level. The close participation of the counselors before, during and following BSEP classes, the "caring" attitude they display and their interest in the soldiers' progress enhances the learning process and hastens achievement of the program's goals.

Results of an average six month BSEP program at a small Education Center in Germany are shown below. A total of 221 soldiers participated in five sessions with twenty classes. There were four BSEP I Communications classes, three BSEP II Reading classes, seven BSEP II English classes and six BSEP II Math classes.

CLASS SUBJECT	NUMBER OF STUDENTS	AVERAGE INCREASE IN GRADE LEVEL		LOWEST POST TEST SCORE		HIGHEST POST TEST SCORE		GREATEST GRADE LEVEL INCREASE	
		Area A	Area B	Area A	Area B	A	B	A	B
		Mech-anics	Spelling	ME	SP	ME	SP	ME	SP
BSEP I COMMUN.	33	2.25	1.0	4.4	3.0	10.1	12.2	7.1	5.7
		ME	SP	ME	SP	ME	SP	ME	SP
BSEP II ENGLISH	95	1.41	.77	5.1	5.1	12.2	12.0	7.0	5.4
		Vocab.	Compre-hension	VOC	COM	VOC	COM	VOC	COM
BSEP II READING	22	1.30	3.17	7.0	6.9	10.0	11.0	2.7	4.0
		Compu-tation	Problems	COMP	PROB	COMP	PROB	COMP	PROB
BSEP II MATH	66	3.3	1.73	7.2	6.4	12.3	12.5	7.1	5.7

While not all soldiers experienced an increase in grade level as demonstrated through TABE results, and while factors such as physical condition and attitude may have been influential in the scores achieved, the above chart does indicate that some form of behavior modification has taken place through the soldier's participation in the BSEP program. The average increase in grade

level that took place over the six month period described in the chart is typical of the program in Europe. The lowest post test scores and highest post test scores shown above are averages of the lowest and highest scores achieved in each series of classes for the subject. The areas of English and Math appear generally to be the areas of greatest remedial need. Discussions with commanders and military supervisors do indicate that there is an improvement in the job performance of soldiers who have taken BSEP classes.

CONCLUSION

The BSEP program was developed to solve a problem, the problem of soldiers with inadequate basic skills to function efficiently in an increasingly sophisticated technological Army. With concentrated instruction in the subject and supportive guidance counselors the problem is being solved. However, without adequate measuring instruments the solution would be much more difficult. The TABE is providing that measuring instrument.

The future role of the TABE and counseling with the TABE appears to lie in the direction of diagnostic testing to predict success in retaking the ASVAB. Each Education Center is, hopefully, in the process of eliminating the need for the BSEP program. A need for further evaluation of TABE results and study to develop coorelations with the ASVAB exists. Whether or not the TABE continues to be utilized as it is now will probably depend on the development of such studies.

PERFORMANCE MOTIVATION

Chair: Glenda Y. Nogami

Task performance is mediated by three broad categories of variables: Characteristics of the individual, characteristics of the task, and intrinsic/extrinsic motivation. The papers presented in this symposium reflect research in these three areas which are being supported throughout the Armed Forces. Robert Philips presented an overview of the National Youth Attitude Tracking research he helped conduct at Ohio State. This was a longitudinal study of the individual characteristics of high school students - the traditional military recruiting pool. Two papers dealt with the interaction of characteristics of individuals and tasks on stress and performance. Siegfried Streufert investigated the effects of information and stimulus load (in a decision making simulation and a video-game task, respectively) on strategic/integrative decision making and planning performance. George Troxler and William Hendrix's paper dealt with stress as a mediating variable for co-worker relations, job enhancement levels, and job satisfaction, with its consequences on job performance. The paper by John O'Hara investigated the differential effectiveness of extrinsic motivators (incentives) for high and low performing recruiters. He suggests some alternative/additional incentives for increasing job performance.

Stress: Its Behavioral and Physiological Consequences

William H. Hendrix, Clemson University
R. George Troxler, School of Aerospace Medicine
Nester K. Ovalle, 2d, Air Force Institute of Technology

Performance of an individual within a job setting can be conceived as an interaction between organizational, non-organizational, and individual characteristics which affect one's productivity. One of the potential effects of these factors is that of distress by the individual which affects his or her performance. Generally, if an individual's stress is continually increased, a point will be reached where performance decreases as stress level increases. Along with this decreased performance for the individual is the likelihood that group performance will also decrease due to the individual's reaction. The individual may not have time to interact properly with group members. This behavior may take the form of developing a short temper or hoarding information needed for task accomplishment. In addition, the individual may develop physical problems which decrease his or her effectiveness on the job, or if severe enough may require hospitalization. The problems of ulcers, high blood pressure, allergies, and coronary heart disease are believed to be in part precipitated by stress.

Various physiological changes occur when one is exposed to a stressful environment. Two blood components affected by stress are cholesterol and cortisol (an adrenal hormone). Friedman and Carroll (1957) examined tax accountants to determine the effects that heavy work load, high level of responsibility, time pressure, conflict, and job-role ambiguity had on cholesterol level. Their results indicated that there was a marked increase in the blood cholesterol level as the tax-filing deadline approached. After the deadline passed, the cholesterol decreased returning to normal within two months. HDL cholesterol, on the other hand, has been indicated as a coronary heart disease reducing factor (Kritchevsky, Paoletti, and Holms, 1978). That is, as HDL cholesterol increases, there is a decreasing probability of developing coronary heart disease.

Cortisol is an adrenal hormone which is secreted into the blood stream. A series of studies (Brown, Schalch, and Reichlin, 1971; Kopin, 1976; Rubin, Rache, Clark, and Arthur, 1970) have indicated that as stress increases there is a resulting increase in the blood cortisol level. In addition, there is some evidence that increased cortisol levels result in increased total cholesterol levels. This relationship suggests that stress may be a factor in the development of coronary heart disease.

Notwithstanding the laudible research efforts on stress in both the behavioral and medical science areas, there is a need for integrative efforts investigating the relationships between organizational/psychological variables, including productivity, physiological dimensions, and stress. In other words, stress research must be performed to incorporate the concerns and knowledge of the physiological and psychological sciences.

This study is a part of a large scale stress research program to establish relationships between individual characteristics, work group and organizational factors, and extra-organizational factors to organizational and individual outcomes. Specifically, the major organizational outcome identified in this study was work group productivity. The individual outcomes were: (a) potential for coronary artery disease, (b) physiological stress, and (c) perceived stress.

Method

A sample of 436 individuals completed the Stress Assessment Package (version 2), and had their blood drawn. Individuals were DoD civilian and military employees located at installations across the United States. Of these, 269 were males and 167 were females. Participation was on a voluntary basis and anonymity was insured by each subject selecting a number which served as their personal identifier known only to them.

Survey Instrument

The Stress Assessment Package (version 2) used for data collection consisted of 160 items of which 130 were primarily 7-point Likert attitudinal scales and 29 were background iinformation items. The Likert attitudinal items were designed to measure organizational variables (e.g., organizational climate, job enrichment, autonomy, role conflict, and goal setting) and personality variables (e.g., Type A Behavior and Locus of Control). The background information items were used to collect data such as sex category, race, and for personal history items such as smoking, dietary fat consumption and jogging experience.

Procedure

The Stress Assessment Package was administered to volunteers en masse at each administration site. After completing the survey, individuals computed their indices on a series of factors such as assertiveness and Locus of Control. An explanation of these factors and how each was related to stress was provided. Individuals desiring to learn of their cholesterol and cortisol levels had their blood drawn. Almost all individuals completing the survey also had their blood drawn (over 90%). The attitudinal items were factors analyzed with 25 orthogonal factors extracted (Table 1). In turn, the dependent variables/factors of (a) perceived stress, (b) physiological stress (cortisol), (c) potential for coronary artery disease (measured by the ratio of total cholestrol divided HDL clolesterol), and (d) perceived work group productivity were regressed using as independent factors those extracted during factor analysis of the Stress Assessment Package.

TABLE 1.
Orthogonal Factors

Factor No	Label
1	Internal/External Locus of Control
2	Type A/B Behavior
3	Perceived Productivity
4	Job Autonomy
5	Planning Time
6	Intergroup Conflict
7	Task Significance
8	Goal Clarity
9	Need for Enrichment
10	Group Goal Setting
11	Problem Solving Participation
12	Job Enhancement
13	Supervision
14	Supervisory Control
15	Micro Supervision
16	General Organizatinal Climate
17	Organizational Control
18	Co-worker Relations
19	Assertiveness
20	Community/Social Activity
21	Family Relations
22	Exercise
23	Job Satisfaction
24	Tolerace for Change
25	Perceived External Stress

Results and Discussion

What organizational, extraorganizational, and individual factors are predictive of perceived organizational stress? Table 2 gives the regression results with organizational stress as the dependent variable and all the factors listed in Table 1 as independent variables.

TABLE 2.
Regression Analysis Results
Dependent Variable: Perceived Organizational Stress
Independent Variables: All Factors

Factor	Label	R^2	Change in R^2	Beta	Significance
17	Organizational Control	.10800		.119	.017
4	Job Autonomy	.16333	.055	-.227	.001
2	Type A/B Behavior	.22233	.059	.152	.002
6	Intergroup Conflict	.24968	.027	.153	.002
1	Locus of Control	.26621	.017	.099	.047
24	Tolerace for Change	.27798	.012	.138	.005
8	Goal Clarity	.28912	.011	-.141	.007
12	Job Enhancement	.29967	.010	.237	.001
23	Job Satisfaction	.31621	.017	-.185	.004

Generally, individuals had higher perceived stress if they were in organizations that had a high degree of control with low autonomy for individuals, high job enhancement levels, poor goal clarity, high intergroup conflict, and low satisfaction. Those individuals who experienced high degrees of stress were those who tended to be type A, external locus of control individuals who scored high on tolernace to change. This last factor is in line with the literature which indicates rigid individuals are stressed less than more flexible individuals.

What organizational, extraorganizational and individual factors are predictive of physical stress, i.e., cortisol? Table 3 summarizes the regression results with cortisol as the dependent variable and all 25 factors as the independent variables.

TABLE 3.
Regression Analysis Results
Dependent Variable: Cortisol
Independent Variables: Factors

Factor No.	Label		Change in 2	Beta	Significance
24	Tolerance For Change	.02487		-.151	.004
21	Family Relations	.04289	.018	.140	.008
18	Coworker Relations	.05457	.012	-.109	.038

As one would expect, the data indicate that individuals have higher cortisol levels if coworker relations are poor. However, the data indicate those with positive family relationships and high tolerance for change also have high cortisol levels. There is no apparent reason for these unanticipated results except that cortisol is very unstable (i.e., is influence significantly by many factors) and the results maybe only chance variation.

What organizational, extraorganizational and individual facets are predictive of CHD potential (i.e., the ratio between total serum cholesterol to HDL cholesterol)? Table 4 summarizes the results of the regression analysis with the ratio of total cholesterol to HDL cholesterol as the dependent variable and the factors identified by factor analysis as the independent variables.

TABLE 4.
Regression Analysis Results
Dependent Variable: Ratio of Total Cholesterol to HDL Cholesterol
Independent Variables: Factors

Factor	Label	R^2	Change in R^2	Beta	Significance
25	Dietary Fat	.01885		.135	.011
4	Job Autonomy	.03240	.014	.152	.007

The data indicate that an individuals ratio increases with increase in dietary fat and with increased job antonomy. Here we have some indication that the responsibility associated with job antonomy may have a physiological effect, specifically an increased ratio.

What organizational, extraorganizational, and individual factors are predictive of perceived productivity? Table 5 summarizes the results of the regression analysis having perceived productivity as the dependent variable and all the other factors as independent variables.

TABLE 5.
Regression Analysis Results
Dependent Variable: Perceived Productivity
Independent Variables: All Other Factors

Factor	Label	R^2	Change R^2	Beta	Significance
14	Supervisory Control	.04106		.166	.002
7	Job Significance	.06284	.022	.133	.011
18	Coworker Relations	.07424	.011	.103	.049

The data general indicate that productivity increase is related to a job that is high in significance, that has good coworker relations, and has a supervisor who controls the work process.

Summary

Overall, these data indicate that stress, potential for developing coronary artery disease, and perceived productivity are dependent on individual, organizational, and extraorganizational factors. In order to provide optimal effectiveness for the organization while providing for a workers satisfaction and encouraging performance motivation one should ensure that a significant job with adequate supervisory control and good co-worker relations is provided. Generally, the data related to stress and potential for coronary artery disease are consistent with organizational health. That is, a healthy organization does not produce a distress, high coronary artery disease potential individual.

References

Brown, G. M., Schalch, D. S., & Reichlin, S. Patterns of growth hormone and cortisol responses to psychological stress in the squirrel monkey. Endocrinology, 1971, 88, 956-963.

Friedman, M., & Carroll, V. Changes in the serum cholesterol and blood clotting time of men subject to cyclic variation of occupational stress. Circulation, 1957, 17, 852-861.

Koplin, I. J. Catecholamines, adrenal hormones, and stress. Hospital Practice, March 1976, 49-55.

Kritchevsky, D., Paoletti, R. & Holms, W. Drugs, lipid metabolism, and atherosclerosis, New York: Plenium Press, 1978.

Rubin, R. T., Rache, R. H., Clark, B. R., & Arthur, R. J. Serum uric acid, cholesterol, and cortisol levels: Interrelationship in normal men under stress. Archives of internal medicine, 1970, 125, 815-819.

PRELIMINARY ASSESSMENT OF THE ARMY'S INCENTIVE PROGRAM FOR RECRUITERS

John W. O'Hara, Paul A. Gade, Timothy W. Elig,
Newell K. Eaton, and Allyn Hertzbach

US Army Research Institute for the Behavioral and Social Sciences
Alexandria, Virginia 22333

Introduction

Increasing recruiter productivity through the use of incentives is a continuing concern of the U.S. Army Recruiting Command. The problem of increasing productivity becomes more crucial as the need for highly qualified recruits increases. Recruiters are now expected to recruit for quality as well as quantity. The specific purpose of this research effort was to assess the research needs and operational problems of the current U.S. Army recruiting incentive awards system.

The current recruiter incentive system can be divided into three components: performance measurement, consequences of performance, and management of the system. Recruiter performance is measured by how well a recruiter meets his or her "mission box" requirement. The mission box requirement is based on army needs for several categories of recruits with emphasis on quality. To satisfy mission box requirements, a recruiter must each month contract specified numbers of individuals in categories based on education, prior service status, gender, and performance on the Armed Services Vocational Aptitude Battery. A variety of recognition awards are given to recruiters for successfully meeting mission box requirements and a variety of corrective actions may follow when recruiters fail to meet these requirements. The management of the current system is accomplished primarily at recruiting command headquarters.

Method

This research was part of a larger data collection effort conducted between August and October, 1981. Recruiters and station commanders were interviewed and surveyed to determine their knowledge of and attitudes about the current incentive awards program. Recruiter attitudes toward the current award system were examined as a function of gender, performance, satisfaction with recruiting, and recognition received from commanders. Recruiter and station commander suggestions concerning changes in performance measurement, consequences of performance (the awards), and system management were examined as well.

The views expressed in this paper are those of the authors and do not necessarily reflect the views of the US Army Research Institute or the Department of the Army.

Survey and Structured Interviews

The survey consisted of a paper and pencil questionaire that solicited information about demographics, productivity, job satisfaction, personality characteristics, and job preferences. The structured interview covered several topics, one of these was recruiter incentives and motivation. The interview questions were essentially the same for recruiters and station commanders. These were open ended questions, with no restriction on the number of responses an individual could give. The interview responses were content analyzed to identify major categories of responses, and the frequency of responses in those categories reported.

Survey and Structured Interview Sample

Recruiters and station commanders were sampled equally from each of the 5 regional recruiting commands. The total sample included 53 station commanders and 103 recruiters.

The 50 stations were divided among 5 ARI interviewers for survey administration. Survey forms and interviews were completed in the recruiting stations during regular working hours. Interviews were conducted in a private location within the station. Participants were promised confidentiality.

Results

Are the Current Awards Effective?

Recruiter attitudes toward the current awards program were examined by asking: "Do the awards available to recruiters motivate you?" The percent of the sample of recruiters responding "yes" and "no" to the question is shown in Table 1 as a function of recruiter gender, productivity in terms of percent of objective acheived, job satisfaction, and certificates of appreciation received from high-level commanders.

Only 27 percent of the sample of females said that they were motivated by the awards compared to 52 percent of the sample of male recruiters, $\chi^2(1)=5.67$, $p=.017$. Clearly, female recruiters feel especially unmotivated by the awards available to recruiters. Since females were represented at a higher percent in the sample than in the actual recruiting force, the total sample was weighted for the proportion of male and female recruiters in the force. Weighted responses for all recruiters were 46.5 percent "yes," 46 percent "no," and 7.5 percent "no response."

Productivity in terms of percent of objective achieved in the last 6 months was supplied by recruiter self-reports on the questionaire portion of the survey. The reported effectiveness of the awards was related to the productivity of recruiters, $\chi^2(2)= 13.39$, $p=.001$. Recruiters who were below average in productivity said they were extremely unmotivated by the awards while those at exactly 100 percent said they were somewhat unmotivated.

High or low job interest was determined from responses to three questions on the questionaire part of the survey. These questions dealt with job importance and job activities. Recruiters who showed high job interest said they were especially motivated by the awards available to them, $\chi^2(1)=13.82$, $p=.0002$. Also, recruiters who received certificates of appreciation or commendation from high-level commanders at an above average rate said they were especially motivated by the awards, $\chi^2(1)=8.93$, $p=.0028$.

The opinions of station commanders about the effectiveness of the awards system were also assessed. They were asked, "Do the awards available to recruiters motivate them?" Responses were 45 percent "yes," 38 percent "no," and 17 percent "no response."

Table 1

Percent of Recruiter Responses to:
"Do the Awards Available to Recruiters Motivate You?"
By Moderating Variables

Moderating Variable	Percent (Frequency)	
	Yes	No
Gender		
Male	52 (32)	48 (30)
Female	27 (9)	73 (25)
Percent of Objective Achieved		
Above 100	62 (25)	38 (15)
100	40 (10)	66 (15)
Below 100	19 (6)	81 (25)
Level of Job Interest		
High	59 (32)	41 (22)
Low	21 (9)	79 (33)
Number of Certificates Received per Year from a DRC or Higher Command		
High	58 (26)	42 (19)
Low	27 (12)	73 (33)

Note: Total N=103, but there were a few omissions in each section of the table.

In summary, the current award system is most likely to be perceived as a source of motivation for recruiters who are male, above average in

productivity and job interest and receive many certificates of appreciation or commendation. It is least likely to be perceived as a source of motivation for recruiters who are female, average to below average in production, below average in job interest, and receive few certificates of appreciation or commendation. The overall interest in the award system was not high.

What Other Incentives M[illegible] [illegible]d to Motivate Recruiters?

Many recruiters a[illegible] [illegible]n commanders listed a variety of potential incentives when they were asked: "What would motivate you to do even better in recruiting?" or "What motivates recruiters?" These potential incentives are shown in Table 2, listed by percent of recruiters giving the response. The frequencies in this table represent relative importance of responses. There appear to be several potential incentives beyond the recognition awards currently used that are meaningful to recruiters and might be used to motivate them.

Table 2

Potential Incentives Identified by Recruiters and Station Commanders

Incentive	Percentage of Recruiters	Percentage of Station Commanders
Awards	---	38
Better pay and benefits	24	15
Time off	23	15
Better opportunity for promotion	13	19
Choice of assignment	7	---
Personal approval and recognition	6	19
	64[a]	60[a]

[a] Percents do not sum to the total because individuals could make more than one response. The total is less than 100 because other types of responses were also given.

How Can System Management and Performance Measurement be Improved?

Recruiters and station commanders were also asked "How can the award system be improved?" Many of the responses dealt with performance measurement and system management.

Recruiters preferred that performance measurement be based on total numbers put in the army rather than the mission box categories. There was concern with aspects of system fairness such as geographical area differences and the difficulty of earning awards. Other suggestions were that the reception of awards should be more prompt, that the system should be explained better, and that the system should not change so often.

Conclusions

While more evidence is needed before causal interpretations of these relationships are possible, some ideas are worth consideration. Low productivity recruiters might be more motivated by the awards if they had a better chance to get them. Recruiters and station commanders commented that the awards are too hard to get. Hamner and Hamner (1976) state that for rewards to work, people should have a chance to succeed. Of course the above must be balanced by the necessity to differentiate rewards based on performance (Hamner, 1974). Nadler and Lawler (1977) state that individuals have expectations that they can accomplish a level of of performance and expectations of outcomes for that level of performance. Individuals would therefore have expectations concerning their chances of getting awards, and those with low expectations might lose their motivation for the awards.

That female recruiters were not as motivated by the awards as males might be further evidence for sex differences in job orientation as reported by Manhardt (1972) and Schuler (1975). These and other researchers have reported that females show greater interest in social aspects of a job while males show greater interest in career objectives of the job. These differences have been questioned by many investigators reporting no sex differences in job orientation such as Voydanoff (1980), but the issue is not yet settled. Awards might be an aspect of career objectives for recruiters, and therefore of greatest interest to males.

Receiving certificates of appreciation or commendation from high-level commanders correlated positively with being motivated by the awards. That certificates of appreciation or commendation used judiciously would motivate is consistent with recruiter and station commander comments that praise and personal recognition are a desired reward.

The direction of causation between job interest and motivation for the awards must be determined. It is not clear whether poor job interest is the cause or result of poor job performance. It is also not clear whether poor job interest is the cause or result of low interest in the current awards program.

The reward preferences expressed by Army recruiters (Table 2) are more similar to those of civilian sales forces than to those of other military personnel, manufacturing personnel, or public sector personnel (Spector, 1982). This suggests we can be more confident in using information from civilian sales incentive programs to develop hypotheses about recruiter incentives.

These survey results provide information concerning which recruiters are most in need of further incentives, and what changes in the incentives or the system of management are preferred by recruiters. The results will be used in the development of an improved incentive system for Army recruiters.

References

Hamner, W. C. Reinforcement theory and contingency management in organizational settings. In Tosi, H. L., & Hamner, W. C. (Eds.). Organizational Behavior and Management: A Contingency Approach. Chicago: St. Clair Press, 1974.

Hamner, W. C., & Hamner, E. P. Behavior modification on the bottom line. Organizational Dynamics, 1976, 4, 3-21.

Manhart, P. J. Job orientation of male and female college graduates in business. Personnel Psychology, 1972, 25, 361-368.

Nadler, D. A., & Lawler, E. E. III. Motivation: A diagnostic approach. In Hackman, J. R., Lawler, E. E., & Porter, L. W. (Eds.). Perspectives on Behavior in Organizations. New York: McGraw-Hill, 1977, 26-36.

Schuler, R. S. Sex, organizational level, and outcome importance: Where the differences are. Personnel Psychology, 1975, 28, 365-375.

Spector, B. I. Toward developing an army recruiter incentive program: Literature review. Unpublished Manunscript, Bethesda, Maryland: Booze, Allen, & Hamilton Inc. August, 1982.

Voydanoff, P. Perceived job characteristics and job satisfaction among men and women. Psychology of Women Quarterly. 1980, 5, 177-185.

AD P000 818

LOAD EFFECTS ON THE USE OF STRATEGY IN MOTIVATED PERSONNEL

Siegfried Streufert

Pennsylvania State University, College of Medicine, Hershey PA 17033

While much research employs the concept of "motivation" as a dependent or independent variable, motivation may be viewed in terms of a mediating variable as well. This is the approach taken in the present paper. The primary concern of this manuscript is with the effects of information load on performance in two quite different tasks. Load is, without question, a potential stressor (Streufert and Schroder, 1965; Streufert and Streufert, 1982). It is now well known that overload may diminish performance. However, underload, i.e., information deprivation, may also impair performance (cf. Streufert and Streufert, 1978 and the extensive research program of Suedfeld and associates, e.g., Suedfeld, 1978). Load and its potential stressor components would likely affect performance to a lesser degree (or in some cases, not at all) if personnel performing a task were not motivated. Lack of motivation would likely have two quite separate (although interactive) effects: (1) information input would not be taken as seriously, thereby diminishing the effective load level, and (2) performance levels which would be relatively low would provide for lesser differences between diverse load effects (a ceiling effect). The motivated person, on the other hand, would likely be eager to consider all relevant information which he or she receives, and would - if able - achieve his or her optimal level of performance where load levels are conducive so that considerable performance decrements can be measured when load levels represent aversive conditions.

To study the effects of load on performance, we should then consider not only load (as the independent variable) and performance (as the dependent variable) but also stress effects (strain, a mediating variable) and finally motivation. In the present research, motivation levels are held relatively constant at high levels (with one exception mentioned below) by providing environments where incentives for doing well are presented, including competitive challenges and financial rewards. Motivation to perform well can then be assumed to be given (and is demonstrated to exist via manipulation check techniques).

The concept of stress cannot be as easily controlled or held constant if one wishes to study load effects. As stated above, load itself is a stressor. Moreover, its effects are not linear. As described long ago in the Yerkes-Dodson law, stress may be experienced at higher levels both at the low and the high end of the load dimension. Optimal stressor effects may be experienced at intermediate levels.

How does load stress relate to performance? This paper will discuss two sets of research efforts: one is concerned with load effects in complex decision making tasks. The other utilizes a much simpler performance setting: a visual-motor task similar to a video game. For the present purposes the interest is in one specific performance measure: strategic planning (measured as the integration of a current action with a planned future action). Both tasks, despite their considerable differences, allow some strategic planning to occur. We may ask to what degree stress experience (at low, moderate and

high levels) induced by diverse information load levels are likely to alter strategic planning across the diverse task conditions. This paper will initially review the data obtained in a research program on a complex decision making task and will then turn to the visual-motor task. Finally, a comparison of the results from these research programs will be made.

Load and Complex Decision Making

With increasing automation, the number of simple decisions which need to be made by organizational, including military, personnel are likely to continue to decline. However, computers are not (at least certainly not yet) able to aid us in making complex decisions in uncertain conditions in response to complex task demands. At best, automation can produce a greater flow of (hopefully more relevant) information. Increased information, however, may imply increased load experienced by the human personnel that must make the final decision. How can this load be dealt with most effectively?

Unfortunately the standard decision making literature is not of much help in answering such questions. Most efforts to describe and predict human decision making processes have been based on providing alternative choices between fixed outcomes with more or less certain implications (or similar sets of relatively "simple" components in the cognitive decision making process and its informational basis).

Complex decision making in the "real world" has rarely corresponded to such models. For example military decision making at command levels necessarily involves degrees of uncertainty which often are not even resolved after a decision has been made. Wohl (1981) has argued this point rather well. Wohl believes that relatively little agreement among researchers has been reached with regard to the decision process. Decision theorists have been prescriptive rather than descriptive (or analytic) in their efforts. Uncertainty has been concerned with decision input, not with the decision making process itself (e.g. Edwards, 1961). On the other hand, military commanders are necessarily concerned with the "creation, evaluation and refinement of hypotheses" with regard to their situation and with options for responses. These processes are not necessarily "rational" in the sense used by standard decision theory (c.f. Janis and Mann, 1977) nor can they be solely determined from the knowledge of information input. Rather a cognitive analysis is needed. Again, following Wohl, when data is of high quality and options can be specified without error, a mapping process can be designed which translates inputs directly into outputs. However, tactical military decision making is generally characterized by data of limited quality and by open-ended or poorly defined options. Rapid hypothesis formation and option processing is consequently needed. Standard decision making approaches do not provide much information about such processes.

Theory (e.g. Streufert and Streufert, 1978) and research (e.g. Streufert, 1970) by Streufert and associates has attempted to explore load effects on complex information processing under conditions of uncertainty which reflect organizational and military environments more appropriately. The missing elements of uncertainty and lack of immediate feedback are provided. A complex, yet experimental simulation technique (c.f. Fromkin and Streufert, 1976) was developed to provide the necessary task environment for the measurement of complex decision making. Data are obtained via statistical analysis of a time/event matrix (c.f., for example, Streufert and Streufert, 1981) which describes the inputs and outputs to and from decision maker(s) over a specified length of time. Data obtained with this procedure have shown high

levels of reliability. Validity has been demonstrated in executive settings. Applications to senior level military decision making processes are under way.

Clearly, the optimal choice for measuring the presence and the degree of stressors (as components of load) is physiological arousal, measured, for example, in terms of delta (elevations of) systolic blood pressure, diastolic blood pressure and heart rate. For the earlier data on load effects in a complex simulated decision making environment which will be reported here, physiological data are not available. Scale responses (manipulation checks), however, indicated that stress was highest at overload levels (e.g. when load reached or exceeded one item of information every two minutes), high at underload (information deprivation) levels, (e.g. when load levels were at or below one item of information every six minutes) and moderate or low when load levels approached one item of information every three minutes[1]. In other words, we may assume that stress in complex simulation tasks is less associated with intermediate load levels than with low or high information load levels.

A number of performance measures were obtained in the simulation task. For the present purpose, the focus is on responses reflecting the utilization of strategy (in the popular meaning of that term), i.e. planning for future actions. Credit for planning activity (decision integration) was given when a decision was made (entirely or in part) as the basis for a future decision of a different kind, assuming that future decision was indeed carried out later on. The number of decision integrations were counted separately for a number of playing periods in the simulation. Each period (in random order from participants to participants) presented information at a different load level (e.g. 2, 5, 8, 10, 12, 15, 25 items of information per 30 minute playing period). Optimum integrative decision making performance i.e. the highest level of strategic planning activity, was obtained at load level 10, that is when one item of information was presented every three minutes. Figure 1 represents a typical relationship between load and mean (across 20 groups of subjects) integrative (strategic planning) performance from one series of experiments (carried out in various countries and with various populations). The data obtained show high levels of reliability across the various settings, samples and experimenters.

For performance in complex simulation experiments, then, it appears that load is associated with stress and that stress, probably in part as a mediator variable, has a direct effect on strategic planning performance. It may be mentioned as an aside, that individual differences in cognitive complexity have considerable modifying effects on the observed load effects on performance: While an inverted U shaped curve is obtained for both more and less cognitively complex persons (as in Fig. 1), the elevation at optimal load levels is considerably higher for the more cognitively complex individuals. Further, other measures of performance (e.g. quantity of decision making output and the number of responses which can be characterized as inappropriate to the task at hand) tend to show a curvilinear rise with increasing load levels. They are, in other words, less affected by underload stress.

---

[1]A current research program will obtain physiological strain measures for a similar data set.

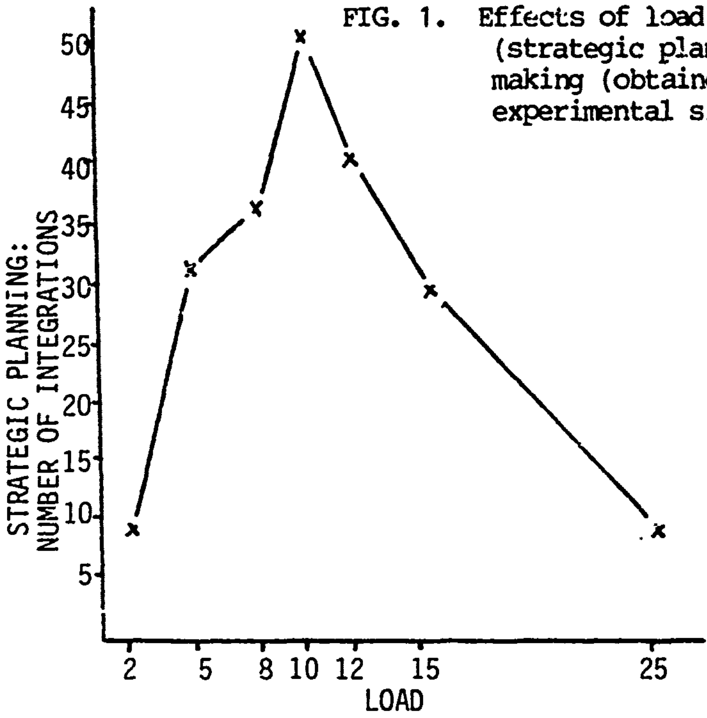


FIG. 1. Effects of load on integrations (strategic planning) in decision making (obtained in a complex experimental simulation) task.

Load in a Visual-Motor Task

For a research project which is presently in progress, a visual-motor task has been developed (e.g. Streufert, Streufert and Denson, 1982). The participant in the task, working individually, is introduced to a video-game type setting. He or she must guide a scoop through a matrix presented on a TV screen, collecting stationary squares within the matrix while avoiding circular objects which move randomly through that matrix. From one through nine circular objects may be presented and scoop and objects may move at various predetermined speeds. The participant in the task should, if possible, avoid moving his or her scoop through any corridor in the matrix more than once: points are lost in traversing blank (empty) spaces where squares were already collected previously. More serious, however, is a collision with any one of the circular objects: a collision results in a vibration of the TV screen, a loud noise, and an instant loss of 100 points.

To obtain as high a score as possible, the participant must not let squares stand in locations where longer empty spaces must be traversed at a later time to collect those squares. While the participant is urged to be as effective as possible to obtain as high a score as possible (very high comparison scores supposedly achieved by others are provided) he or she is not told what the best strategy for achieving high scores would be. Load in this task is represented by the number of circular objects with which the participant has to deal. Strategic planning is scored in terms of the number of times a nearby (but not in direct line) square is picked up (positive score) and the number of times the participant fails to make a turn in the matrix which would have provided less costly access to squares later in the task (negative score).

Stress in this task was measured as physiological strain experienced during task performance. Systolic, diastolic blood pressure and heart rate was obtained in intervals of two minutes during all task periods. Following a

warm-up trial period (with low speed and only one circular object present), participants worked to erase all squares in the matrix during four additional playing periods. They experienced (in randomized sequence) either 2, 4, 6, or 8 circular objects. Speed during these periods was moderate. The data indicate that load resulted in a linear decrease in systolic blood pressure and heart rate. At the same time, a linear increase in diastolic blood pressure was obtained. Diastolic blood pressure is associated with peripheral constriction any may represent the measured equivalent of nor-epinephrin showers into the bloodstream. In this task, then, increasing load is associated with increasing strain. Manipulation check scale responses collected after each task period confirm that participants felt increasingly stressed as load increased.

Performance (strategy) was inversely related to the strain measure. As load was (randomly) increased, strategy scores decreased. For the higher load levels (6 and 8 circular objects in the matrix), the obtained strategy scores fell to levels below zero, in other words, strategic errors exceeded positive strategy actions. The data are shown in Figure 2. As an aside, it may again be mentioned that other performance scores (in addition to the strategy measure) were obtained as well. Total score (the number of points credited for collecting squares minus points for empty spaces traversed and minus 100 points for each collision) showed a similar effect as did strategy. Risk taking, on the other hand, showed a linear increase with increasing load.

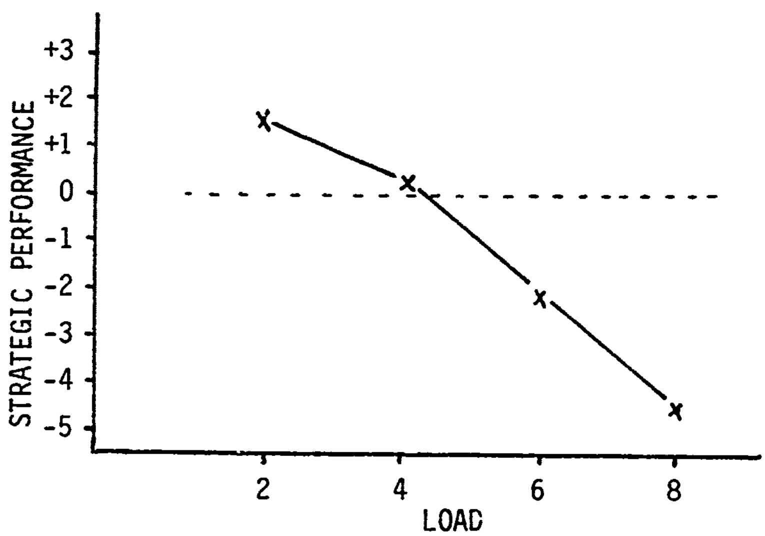


FIG. 2. Effects of load on strategic performance (planning) from a visual-motor task.

Load Effects Across Tasks

The data obtained in both research settings indicate that load does affect strategic performance. It should be remembered, however, that the participants were highly motivated, and that equivalent effects may not be expected in less or in unmotivated persons. For that matter, one study in which motivation was diminished through an experimental manipulation (utilizing the complex simulation task) produced considerably diminished load effects.

Particularly interesting is the reliable association of load effects with perceived stress and/or physiological strain and with strategic performance. Both related to load levels as U shaped vs. inverted U shaped curves for the

complex task. Both showed a linear (rising vs. declining) function for the simpler visual-motor task. It then appears likely, that at least strategic (planning) performance due to load in motivated personnel may be mediated directly by strain, i.e. stress experience. Providing training, or making (where possible) changes in the task environment to decrease stress while maintaining motivation may well aid in assuring higher task performance (at least in strategic planning activities) across diverse task settings.

References

Edwards, W. Behavioral decision theory. Annual Review of Psychology, 1961, 12, 473-498.

Fromkin, H.L. and Streufert, S. Laboratory experimentation. In M. Dunnette, (Ed.): Handbook of industrial and organizational psychology. Chicago: Rand-McNally, 1976.

Janis, I.L. and Mann, L. Decision making New York: Free Press, 1977.

Streufert, S. Complexity and complex decision making. Journal of Experimental Social Psychology, 1970, 6, 494-509.

Streufert, S. and Schroder, H.M. Conceptual structure, environmental complexity and task performance. Journal of Experimental Research in Personality, 1965, 1, 132-137.

Streufert, S. and Streufert, S.C. Behavior in the complex environment Washington, D.C.: Victor H. Winston and Sons and John Wiley, 1978.

Streufert, S. and Streufert, S.C. Stress and the measurement of task performance: I. Decision making in complex tasks. Technical Report ONR No. 3, 1981.

Streufert, S. and Streufert, S.C. Effects of load stressors on performance in a multidimensional visual-motor task. Technical Report ONR No. 6, 1982.

Streufert, S., Streufert, S.C. and Denson, A.L. Information load stress, risk taking and physiological responsivity in a visual-motor task. Technical Report ONR No. 8, 1982.

Suedfeld, P. Characteristics of decision making as a function of the environment. In B. King, S. Streufert and F. Fiedler (Eds.): Managerial control and organizational democracy. Washington, D.C.: Victor H. Winston & Sons and John Wiley, 1978.

Wohl, J.G. Force management decision requirements for air force tactical command and control. IEEE Transactions on Systems Man and Cybernetics, 1981, SMC11 618-639.

ASSESSMENT OF PRACTICE EFFECTS: ASVAB & PACE

Chair: Hilda Wing

Historically, the effects of practice on standardized tests have much anecdotal but little systematic evidence. If a test or test part be subject to practice effects, reliability and predictive validity may be impaired. A special concern is the particular practice effect available to a few examinees via coaching or breaches of security. This symposium presented the results of four empirical studies of two nationally administered standardized multiple abilities test batteries: ASVAB and PACE. The data answer some questions, clarify others, and expose new concerns.

AD P000819

THE EFFECTS OF PRACTICE ON THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY

Bruce K. McCormick & William P. Dunlap
Tulane University
New Orleans, LA

Robert S. Kennedy
Canyon Research Group
Orlando, FL

&

Marshall B. Jones
Hershey Medical Center
Pennsylvania State University
Hershey, PA

28 October 1982

ABSTRACT

The Armed Services Vocational Aptitude Battery (ASVAB) was administered five separate times to fifty-seven men and women of military service age. The objective was to determine to what extent means and cross-session correlations are stable over sessions. Ten individual subtests, the derived ASVAB area composites (N=10) and the Armed Forces Qualification Test (AFQT) were examined for stability. The means and dispersions of scores for this population were below the national average. Means increased over sessions .5 standard deviations or more on half the subtests and consequently on most of the composite scores. Correlations for the composites were largely stable over sessions. Correlations between composites were generally lower than within composites. The implications of practice effects for paper and pencil as well as automated selection tests are discussed.

Opinions or conclusions contained in this report are those of the authors and do not necessarily reflect the views or the endorsement of the Department of the Army. This research was supported by the United States Army Research Institute under the supervision of Dr. Hilda Wing (MDA 903-82-M-3943).

INTRODUCTION

Although psychologists have known since at least 1920 that mental test scores frequently increase with practice (Dunlop & Snyder, 1920; Gundlach, 1926; Thorndike, 1922), few studies involving multiple testing have been conducted. In recent years there has been an increased interest in practice and coaching effects (Anastasi, 1981; Catron & Thompson, 1979; Messick & Jungblut, 1981; Whimbey, Carmichael, Jones, Hunter & Vincent, 1980 Wing, 1980), but few studies have been conducted which involve more than two or three replications. What evidence there is, however, suggests that repeated testing may produce appreciable effects on the mean. Mackman, Bittner, Harbeson, Kennedy and Stone (1982) found that inter-session correlations on the Wonderlic were stable over 18 replications but the means increased over 21 percentile points, suggesting that exposure history would be an important variable, were one to employ this test in the assignment of personnel. The Wonderlic possesses many of the same item types as the Armed Services Vocational Aptitude Battery (ASVAB). The purpose of this study was to determine to what extent means and cross-session correlations of the ASVAB are stable over sessions.

METHOD

Subjects: The subjects for this study were 57 men and women enrolled as trainees in the Job Corps Center, Shreveport, LA. Thirty-four subjects were male (29 Black and 5 White) and 23 were female (19 Black and 4 White). It was explained that subjects would be required to take the ASVAB on five consecutive mornings and that the results would be used for research purposes. Additionally, trainees were told that their scores from the first day of testing could be used for determining their eligibility for enlistment in the armed services, if they so desired. It was emphasized that participation in this project would not obligate subjects to consideration for military service. Trainees were also told that they would be paid for their participation contingent upon completion of all five days of testing. The first 60 volunteers were selected. On the second day of testing two subjects dropped out of the study and a third quit on the fourth day. All three who left quit due to unforeseen work, school or family circumstances.

Apparatus and Procedure: Five forms of the ASVAB were administered from 8:00 AM to 12:00 noon in a group setting for five consecutive days. On each day of testing all subjects took the same form of the ASVAB. The order of administration was: Form A, B1, B2, C1, C2. Forms of the ASVAB having the same letter designation also had identical items comprising the subtests of General Science (GS), Coding Speed (CS), Auto & Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI). Paragraph Comprehension (PC), Arithmetic Reasoning (AR), Numerical Operations (NO), and Word Knowledge (WK), were different across forms. (These are described in greater detail elsewhere in Ree, Mullins, Mathews & Massey, 1982 and Kass, Mitchell, Grafton & Wing, 1982.) Administration followed standard procedures and was conducted by members of the Shreveport Military Enlistment Processing Station (MEPS). Neither coaching nor feedback was given to subjects during the days of testing.

Scoring: Subjects' responses were made on answer sheets which were scored by computer at the MEPS on the afternoon of each day of the project. Results for the ASVAB subtests were combined to form composite scores for AFQT and for the ten aptitude areas.

RESULTS

Means: Significant linear trend, indicating an improvement with practice in the absence of feedback, occurred with four test sections: CS, NO, MK, and MC (Note 1). The means and associated p-values are presented in Table A. The most dramatic increases were for CS and NO, where the fifth test performance exceeded the first test performance by 48.3% and 27.0%, respectively. No test showed a signficant drop with practice. However, both WK and PC showed significant quadratic (U-shaped) changes with administrations, which suggests possible motivational deficits on the intermediate Days 2, 3, and 4. The significant quadratic component for CS was apparently due to the rapid increase in mean score from Day 1 to Day 2, followed by a slower increase thereafter. These mean scores on the first administration are slightly more than one standard deviation below those reported in two reference samples (Kass et al., 1982 and Ree et al, 1982), but for those tests which showed improvement (viz., CS, NO, MK, MC) somewhat less than a standard deviation in later sessions. The standard deviations were constant over sessions and about 75% of the size of the reference samples.

Significant linear trend occured for all the composites except General Technical (GT) and Skilled Technical (ST) (Table B). In the case of General Maintenance (GM), and Electronics Repair (EL), the increase while significant was small (< .2 standard deviations). Composite score group means were approximately half a standard score less in this sample than in a large reference population (Kass et al, 1982). In the first session the average composite score one standard deviation above the mean was 76.2. After five sessions it was 80.6 ($p<.001$).

To better study to what extent practice might help this below-average group, a fine grained analysis of the EL and CO composite was performed. Only five subjects consistently scored higher than 85 on EL, and these same five (and none other) scored higher than 85 on CO. Persons whose scores were slightly below 85 (e.g., 77+) on their first or second administration tended toward higher scores later in practice, but these changes were neither dramatic nor consistent and rarely would they have been sufficient to be able to influence an administrative decision. Individual scores of Surveillance/Communications (SC) (which showed the greatest relative improvement over sessions), revealed no important differences from those seen with EL and CO. Similar relations were also seen for the other composite scores: Field Artillery (FA), Operator/Food (OF), Mechanical Maintenance (MM), Clerical (CL) and (ST).

On the first administration 31 subjects obtained scores of less than 11 for the AFQT. Of these subjects, eleven later achieved higher than 11 at least once and five of them did so on two or more occasions.

Correlations: The intercorrelations across five repeated administrations of each subtest of the ASVAB are presented in Table C. For each subtest a single factor accounts for the bulk of the covariance between sessions. In this sample, eight of the ten subtests either appear to improve with practice or to stay the same. The between session composite correlations are all greater than r = .70 and appear constant or to increase slightly with practice.

Factor Analysis: When five administrations of all ten composite scores were cast into a single 50-variable matrix, the eigenvalue of the first principal factor was 35.1 while the second eigenroot was 4.2 and the third, 1.4; all the rest were approximately equal to or less than unity. For these below-average performers, a single factor accounts for the bulk of the common variance among the area composites.

DISCUSSION

In the present sample differential stabilization (Jones, 1981) with practice is not a problem in ASVAB. All ten subtests are more/less differentially stable on the first administration and remain so. The same is true for the ten aptitude area composites. In neither the subtests nor the area composites is there any appreciable differential change with practice.

Mean changes present more of a problem. Four of the subtests show significant increasing linear trend with practice, and four of the area composites show increases from the first to the fifth administration of .5 standard deviations or more. These changes are sufficient to warrant some concern, although they are not surprising in light of the Mackaman et al (1982) finding of almost 21 percentile points improvement with practice in a population whose mean score began at the 50th percentile.

Several of the correlations for aptitude area composites tend to increase with practice, a finding which has been reported many times before in repeated measures testing (cf. Kennedy, Bittner, Carter, Krause, Harbeson, McCafferty, Pepper & Wiker, 1981). Whether this is due to the restricted range of the present sample or is a more generalizable finding awaits further study. It would appear advisable to attempt to replicate this outcome in a larger and more representative population. Study in a more heterogeneous sample might also reveal that several test administrations would provide a more accurate assessment of an individual's aptitude. It would be useful to study whether certain persons might profit better than others by extra test taking. It is possible that more accurate classification of low-scoring applicants into suitable MOS could be made with repeated measurements.

If test automation of ASVAB proceeds further, it may be helpful to study practice effects. This helpfulness depends on exploiting the possibilities of the new technology by developing new tests, tests that involve more elements of a perceptual, information processing, psychomotor and decision-making sort. It is offered that microcomputer video games might provide a fertile target of opportunity (Jones, Kennedy & Bittner, 1981). It should be noted that when automated, these and other such tests usually involve implicit knowledge of results, which might be expected to

show greater changes in the mean than were found in the present study. Consequently it is likely that with practice they will show appreciable differential change (Jones, 1981) as well. The most promising possibility of introducing more heterogeneity into the ASVAB will also probably revive stabilization with practice as a major concern.

References

Anastasi, A. Coaching, test sophistication, and developed abilities. American Psychologist, 1981, 36, 1086-1093.

Catron, D. W. & Thompson, C. C. Test-retest gains in WAIS scores after four retest intervals. Journal of Clinical Psychology, 1979, 127 352-357.

Dunlap, K. & Snyder, A. Practice effects in intelligence tests. Journal of Experimental Psychology, 1920, 3, 396-403.

Gundlach, R. The effects of practice on the correlations of three mental tests. The Journal of Educational Psychology, 1926, 127, 387-400.

Jones, M. B. Differential process in acquisition. In E. A. Bilodeau & I. McD. Bilodeau (Eds.). Principles of skill acquisition. New York: Academic Press, 1969, 141-170.

Jones, M. B., Kennedy, R. S. & Bittner, A. C. Jr. A video game for performance testing. American Journal of Psychology, 1981, 94, 143-152.

Kass, R. A., Mitchell, K., Grafton, F. & Wing, H. Factor structure of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, & 10: Army applicant sample. Submitted for publication July 1982.

Kennedy, R. S., Bittner, A. C., Jr., Carter, R. C., Krause, M., Harbeson, M. M., McCafferty, D. B., Pepper, R. L. & Wiker, S. F. Perfomance Evaluation Tests for Environmental Research (PETER): Collected papers. Naval Biodynamics Laboratory Report No. 80R008, New Orleans, LA, July 1981.

Mackaman, S. L., Bittner, A. C., Jr., Harbeson, M. M., Kennedy, R. S. & Stone, D. A. Performance Evaluation Tests for Environmental Research (PETER): Wonderlic Personnel Test. Psychological Reports, 1982, in press.

Messick S. & Jungblut, A. Time and method in coaching for the SAT. Psychological Bulletin, 1981, 89, 191-216.

Ree, M. J., Mullins, C. J., Mathews, J. J. & Massey, R. H. Armed Forces Vocational Aptitude Battery: Item and factor analysis of Forms 8, 9, and 10. Report No. 81-55. USAF Human Resources Laboratory, Brooks AFB, TX, March 1982.

Thompson, C. C. The effects of practice on intelligence tests. Unpublished manuscript, Wake Forest University, 1973.

Thorndike, E. L. Practice effects in intelligence tests. Journal of Experimental Psychology, 1922, 5, 101-107.

Whimbey, A., Carmichael, J. W., Jones, L. W., Hunter, J. T. & Vincent, H. A. Teaching critical reading and analytical reasoning in Project Soar. Journal of Reading, October 1980, 6-9.

Wing, H. Practice effects with traditional mental test items. Applied Psychological Measurement, 1980, 4, 141-155.

Note 1 - One or more of these subtests comprises 9 out of 10 composites -- the exception being GT.

APPENDIX I

TABLE A. MEANS FOR ALL SUBJECTS FOR FIVE SUCCESSIVE TEST ADMINISTRATIONS ORDERED BY STRENGTH OF LINEAR TREND WITH LINEAR AND QUADRATIC PROBABILITIES

Section	A	B1	B2	C1	C2	Linear	Quad
Coding Speed	26.7	33.6	36.0	34.9	39.6	.0000	.0081
Numerical Oper	24.1	26.4	26.4	29.4	30.6	.0000	.7602
Math Know	6.8	6.7	7.2	8.4	7.7	.0017	.7758
Mech Comp	8.1	7.7	7.6	8.7	8.5	.0200	.1316
Auto & Shop Info	7.3	7.6	7.8	7.8	8.1	.0757	.6083
Gen Science	8.6	8.1	7.9	8.4	8.1	.0824	.2023
Word Know	13.0	12.7	12.4	10.7	13.1	.1698	.0041
Electronics Info	6.5	6.3	6.7	6.8	7.0	.2280	.4727
Arithmetic Reas	8.9	8.4	10.0	9.4	9.3	.2730	.3533
Paragraph Comp	6.6	4.9	5.6	4.8	6.7	.5982	.0001

TABLE B. TREND OF MEANS AND STANDARD DEVIATIONS FOR 10 AREA COMPOSITES OVER 5 ADMINISTRATIONS

Means

		A	B1	B2	C1	C2	Linear	Quad.
1	GM	64.9	65.1	65.1	67.0	66.9	.0112	.6220
2	EL	66.2	67.0	67.3	69.9	68.0	.0152	.5240
3	CL	69.3	72.5	72.8	73.4	79.0	.0000	.0752
4	MM	65.2	66.2	67.0	70.2	70.8	.0000	.6040
5	SC	66.5	68.9	70.3	70.0	74.5	.0000	.3168
6	CO	66.2	69.1	70.3	70.7	72.4	.0000	.2456
7	FA	68.6	72.2	73.4	75.6	76.1	.0000	.0890
8	OF	64.9	64.7	65.8	66.8	70.0	.0000	.0080
9	ST	66.0	63.0	63.5	64.7	66.5	.2034	.0010
10	GT	67.6	65.6	67.7	63.2	68.0	.5655	.0202

Standard Deviations

		A	B1	B2	C1	C2
1	GM	11.5	11.8	11.8	12.7	12.7
2	EL	12.4	12.7	11.8	12.9	13.5
3	CL	13.9	15.4	16.6	15.6	16.5
4	MM	11.4	11.4	11.3	12.2	11.7
5	SC	11.6	12.0	14.0	13.7	13.6
6	CO	12.4	9.9	10.2	11.6	11.9
7	FA	12.3	11.0	11.8	11.7	12.4
8	OF	11.5	11.2	11.8	11.7	12.4
9	ST	10.6	11.7	13.4	11.0	12.6
10	GT	12.8	12.9	13.2	13.1	13.5

TABLE C. INTER-ADMINISTRATION CORRELATIONS OF THE TEN TEST SCORES

General Science

	2	3	4	5
1	.66	.68	.72	.72
2		.74	.73	.70
3			.76	.79
4				.82

Arithmetic Reasoning

	2	3	4	5
1	.46	.56	.71	.63
2		.54	.51	.65
3			.58	.70
4				.73

Word Knowledge

	2	3	4	5
1	.70	.73	.67	.79
2		.80	.71	.80
3			.78	.83
4				.77

Paragraph Comprehension

	2	3	4	5
1	.62	.59	.58	.58
2		.69	.47	.69
3			.60	.66
4				.57

Numerical Operation

	2	3	4	5
1	.85	.86	.85	.86
2		.90	.90	.87
3			.90	.86
4				.93

Coding Speed

	2	3	4	5
1	.80	.73	.73	.67
2		.86	.82	.77
3			.85	.80
4				.86

Auto & Shop Information

	2	3	4	5
1	.44	.45	.66	.67
2		.58	.41	.39
3			.47	.54
4				.72

Mathematical Knowledge

	2	3	4	5
1	.24	.46	.52	.50
2		.39	.36	.26
3			.54	.42
4				.46

Mechanical Comprehension

	2	3	4	5
1	.20	.58	.38	.46
2		.48	.45	.40
3			.57	.56
4				.76

Electronics Information

	2	3	4	5
1	.64	.51	.54	.52
2		.54	.61	.54
3			.43	.35
4				.70

TABLE D. VARIMAX-ROTATED FACTOR LOADINGS FOR THE TEN TEST SECTIONS AS FUNCTIONS OF TEST ADMINISTRATION NUMBER

		Factor 1 Administration					Factor 2 Administration					Factor 3 Administration				
	Section	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	Gen Sci	.78	.68	.80	.77	.78	-	-	-	-	-	-	-	-	-	-
2	Arith Reas	-	-	-	-	-	-	-	-	-	-	.66	.69	.76	.66	.75
3	Word Know	.73	.68	.70	.52	.70	-	-	-	-	-	-	-	-	-	-
4	Para Comp	.53	.60	.55	-	-	.54	-	.57	-	.57	-	-	-	-	-
5	Num Oper	-	-	-	-	-	.76	.80	.80	.84	.78	-	-	-	-	-
6	Code Speed	-	-	-	-	-	.78	.86	.89	.85	.85	-	-	-	-	-
7	Auto&Shop	.65	.69	.66	.69	.75	-	-	-	-	-	-	-	-	-	-
8	Math Know	-	-	-	-	-	-	-	-	-	-	.63	.63	.55	.70	.56
9	Mech Comp	-	-	-	.65	.69	-	-	-	-	-	-	-	-	-	-
10	Elec Inf	.73	.73	.62	.72	.67	-	-	-	-	-	-	-	-	-	-

ALTERNATIVES TO THE ALL-VOLUNTEER FORCE?

Chair: Michael Berger

The draft ended in 1973, and the All-Volunteer Force (AVF) was established. Selective Service registration ended in 1975, but was resumed in the summer of 1980, in response to dangers (such as the invasion of Afghanistan) which suggested the need for a more rapid manpower mobilization capability if the nation ever faced a military threat. This panel addressed the pros and cons of the AVF and its impact on military readiness.

AD P000820

IS IT NECESSARY TO SEEK AN ALTERNATIVE TO THE AVF?

B. Michael Berger

Deputy Manager, Analysis Division

National Headquarters, Selective Service System

This afternoon we will hold a panel discussion on the need for seeking an alternative to the All Volunteer Force. The idea for this panel developed this past Spring, after I read an article by Harvard University doctoral candidate Eliot Cohen which appeared in the April 1982 issue of Commentary magazine. Cohen's article, "Why We Need a Draft," examined the purported advantages of an All Volunteer Force, and compared them with his concept of the problems associated with continuing the all volunteer policy. He concluded that there was a need to return to a limited or full-scale induction process (Cohen).

Let me summarize the events of the past several years as background for today's discussion. As you know, the military draft ended in 1973, when President Nixon permitted the Congressional authority to induct men into the armed forces to expire. President Nixon based his decision to end the draft, in part, on recommendations of the Gates Commission, a panel appointed by him to study the feasibility of an all volunteer military. The All Volunteer Force was created in concert with the Commission's recommendation that it be backed up by an ongoing registration process. The Selective Service System continued registering men until 1975, when the process was ended by President Ford. Selective Service then went into "deep standby" and remained in that posture until late 1979 when President Carter ordered the resumption of registration and revitalization of the Selective Service in response to dangers (such as the invasion of Afghanistan by Soviet forces) which suggested the need for a more rapid manpower mobilization capability if the nation ever faced a military threat. In January 1982, President Reagan reaffirmed the need for continuing peacetime registration as assuring overall preparedness. He made it clear, however, that he would stick with the All Volunteer Force and could not forsee a need for the resumption of the draft. Since mid-1980, more than 8.5 million men have been registered, and overall registration compliance stands at better than 94 percent.

The nine year AVF experience has been coupled with dramatic changes in the operation and management of the armed forces of the United States. Wages have been substantially increased to attract and retain quality personnel; training has been redesigned to insure the mastery of basic and job related skills; the role of women in the forces has been expanded to include their integration into the military academies and a wide variety of job fields; there has been a strong effort to insure equal opportunity for men and women almost everywhere in the forces, and enlistment standards have been changed to emphasize high school graduation as a "minimum" prerequisite. As recently as this past October, the Military Manpower Task Force, established by President Reagan and chaired by Defense Secretary Caspar Weinberger, reported that it is likely that the armed forces can achieve their goal of growing by 188,000 men (and women) over the next five years without resorting to a draft, provided that military pay keeps pace with wages in the civilian sector. The task force noted the continuing rise in percentages of recruits scoring above national averages on the Armed Forces Qualification Test. Secretary Weinberger, in response to a comment that the depressed economy was a major factor behind improved recruitment, noted that it was only "one factor." He contended

that the rise in enlistments and reenlistments was due also to the fact that "it is again an honor to wear the uniform"...that, "There has been quite a change in the way the military is viewed." (Washington Post).

In spite of reports of success in the AVF, the program is not without its critics. Some who challenge the AVF concept suggest that the military is attracting a less educated class of soldiers and that military standards have been changed to make these soldiers appear better qualified than they really are. It is suggested that training has been simplified to compensate for the lack of qualification, and that too many soldiers continue to fail when faced with tests of the most basic skills. It has been suggested that the Pentagon may be deluding itself into believing that the high school diploma, long the "standard" of educational achievement for the enlisted forces, reflects real academic performance and perseverance, especially when well over three-quarters of American youth now complete high school. There have been suggestions that the number of women in the armed forces and their training and job assignments may be degrading the overall capability to fight. Others suggest that the spillover of the women's rights movement into the military is just another manifestation of the permissiveness sweeping the nation. And, naturally, there is a chorous of voices crying that illness in the economy is the only thing keeping the AVF together. In essence, critics suggest that the AVF is a failure which should, as quickly as possible, be replaced by a draft or other manpower procurement alternative which will raise the qualifications and capabilities of the military.

These conflicting views on the concept of the All Volunteer Force will form the basis for today's discussion. Panel members will address the changes which have occurred in the armed force since implementation of the All Volunteer Force, and consider, as they deem appropriate, the issues I have described. They will endeavor to analyze conditions in the force, discuss advantages and disadvantages of continuing the AVF, and consider whether an alternate form of manpower procurement is in fact necessary. Discussion will focus on the types of persons being attracted to the AVF and consider their motivations and qualifications. The role of women will be considered in terms of their impact on war fighting capabilities.

Each member of the panel will have the opportunity to present his or her views on the topic and issues. We will then have a free discussion of the issues amongst the panel members. Following a break we will open the discussion to members of the audience, first by responding to written questions submitted during the break, then to spontaneous questions from the floor. We hope today's program proves interesting and informative.

This paper represents the views of the presenter. It has not been endorsed or rejected by the Selective Service System, and is not an Agency position.

References

Cohen, Eliot A., Why We Need a Draft. Commentary, April 1982.

Voth, Harold M., The Future of America. Reprint from Military Medicine, Vol. 145, No. 3, March 1980.

The Washington Post, Military Held Likely to Meet Manpower Goal Without a Draft, October 19, 1982.

AD P000821

CITIZENSHIP AND MILITARY SERVICE IN AMERICA

David R. Segal

Department of Sociology, University of Maryland
and
The Twentieth Century Fund

THE TRIPOD OF COMBAT EFFECTIVENESS

The literature on military performance that has accumulated over the past century suggests that the combat effectiveness of soldiers stands on the tripod of cognitive ability, cohesion, and citizenship. I am not suggesting that the three legs of this tripod have received equal attention in the literature, or that they are equally important. They have not, and they probably are not. The first has received continuous scrutiny in the American forces during this century. The second is now experiencing a renaissance after becoming dormant in the post-World War II years. The third is closer to extinction than to dormancy, and is barely admissable as a topic in polite conversation. That we cannot estimate the relative importance of these three components forces us to confront the possibility that the differentials in the attention that we have payed to them may be counter-productive. American military manpower and personnel policies have emphasized the first, have only recently begun to attend to the second, and have studiously avoided the third. If cognitive ability is not, by a very wide margin, the most important of the three, than current policy is not optimally supporting combat effectiveness. The recent improvement in accession quality in the American forces is not due to the policy of maintaining an all-volunteer force, but rather to the youth unemployment rate, and to accession standards that were adjusted in response to that rate. An improvement in the nation's economic health can rapidly lead to a deterioration in the intellectual quality of the armed forces. As a citizen, I get no great feeling of security from the knowledge that the effectiveness of the armed forces that protect me is dependent upon the continued illness of the economy.

COGNITIVE ABILITY

All three legs of the tripod are subject to strengthening or weakening through policy changes. The level of cognitive ability of our military personnel has been shown to be responsive to gross accession strategies (e.g. conscription versus an all volunteer force), and to specific accession standards, as well as to economic factors. An economy in disarray, and the high rates of youth unemployment that it produced, made the all volunteer force a success when it was born in 1973, and saved it in the early 1980s. Although the net effect of psychological

screening of potential military personnel has been called into question (Ginzberg et al., 1959; Janowitz, 1982), a long history of research, in the United States and in other nations, has demonstrated the importance of the "Gideon Criterion" (Wallace, 1982). Smarter soldiers tend to be better soldiers, ceteris paribus (Toomipuu, 1981). Thus, an accession policy that allows the armed forces to draft college educated, college oriented, and college qualified personnel will produce a more effective force than a policy that does not, other things being equal. Similarly, if we assume a direct and strong positive relationship between the amount that we pay our military personnel and their intellectual quality, then a generous compensation system will produce a better force than a stingy one.

COHESION

For much of the nine year history of the all-volunteer force, the American armed services, and particularly the Army, have attempted to substitute manpower policy for personnel policy. That is, they have assumed that by bringing quality personnel into the force, they would produce a quality force. Military analysts have known for a century that combat behavior is influenced by a soldier's peers (du Picq, 1958). The lesson was reinforced by social research on both the American forces (Stouffer et al., 1949; Marshall, 1947) and the Wehrmacht (Shils and Janowitz, 1948) in World War II. The cohesive units called for in the research, however, are not achievable through manpower policy. Accession practices have little impact on unit cohesion. Personnel policies, which dictate how people will be managed after they are in the force, do have an impact on cohesion, but until very recently, personnel policies were not aimed at this goal.

There are reasons why personnel policies have been driven by considerations of efficiency rather than cohesion. Advocacy for the collectivistic orientation in which the concept of cohesion is imbedded came most strongly from those disciplines that have kept closest to their academic roots: history and sociology. The indivualistic orientation that has most influenced policy, by contrast, has been influenced by those disciplines that recognized as legitimate the non-academic employment of their practitioners early: psychology and economics (Segal, 1982). Thus, the sociologists who, having been mobilized for the war effort in the 1930s were demobilized after the war returned to their universities, while psychologists have had a continuing presence in the defense establishment. The ascent of the economists, on the other hand, can be attributed to the decline of the mass armed force. The technologies of air power and nuclear weaponry that burst forth in World War II made obsolete the concepts of widespread demobilization in inter-war periods, and massive mobilization for war, because these new technologies cost us the luxuries of time and distance from the battlefield that had made the mobilization model possible. We moved into an

era in which a large force in being had to be maintained during peace-time (Bachman, Blair and Segal, 1977: 6-9). While Americans were not greatly concerned about defense expenditures during the waging of a popularly supported World War, or during the Cold War period, when there was a widely shared perception of an external threat, it was another matter to maintain high defense expenditures during a period of peace and detente, so efficiency in expenditures had to be demonstrated.

I am not suggesting that cohesion was universally seen as the most important element in combat effectiveness in World War II. There were dissenting voices (see Moskos, 1976: 62). Neither am I suggesting that the importance of cohesion went thoroughly unrecognized for decades. The Army launched a series of experiments in unit rotation to achieve cohesion in the post-World War II years (Segal and Segal, 1984), and the field of experimental small group research would not have been developed had it not been for support from the Navy and Air Force (Segal, 1982). Nor do I want to suggest that current initiatives to achieve unit cohesion are necessarily going to improve the combat effectiveness of the Army. These initiatives are aimed at building primary groups supportive of the Army mission within Army units. This was the model that worked in World War II. Times have changed however. The informal social structure of American society is today characterized better by diffuse social networks than by primary groups, and the maneuver units that host carefully nurtured primary groups in the Army may not be able to maneuver as units on the dispersed battlefield of the future. It may be that military cohesion in today's Army should be structured as a series of social networks rather than primary groups. This may be an area in which we are again preparing to fight World War II. The important point is that the Army's new manning system, initiated by the Chief of Staff and Deputy Chief of Staff for Personnel, Gen. Meyer and Lt. Gen. Thurman, reflects a recognition that accession policies that bring quality people into the service is essential, but is not sufficient. Those people must be shaped into cohesive combat elements if they are to function effectively on the battlefield of the future.

CITIZENSHIP

The history of American military manpower policy is replete with manifestations that as a nation, we have always regarded military service as an obligation and right of citizenship (see Janowitz, 1975: 435). This was reflected in the early militia acts, as well as in the definition of conscripts and of reservists as "citizen-soldiers." Such a conception, however, has a political dimension to it, for citizenship is a political concept, and the trend in recent decades has been toward depoliticization of the American military.

The research on cohesion done in World War II was in fact a major factor in this depoliticization. Janowitz

(1982: 511) feels that the morale studies conducted by Stouffer and his research team (1949) gave military leaders a justification for limiting the political content of their training. Certainly social scientists saw it in this light. Those who were willing to look beyond cognitive factors in seeking explanations of combat effectiveness saw the World War II research on cohesion as unicausal and deterministic, emphasizing primary group formation and denying the importance of other factors such as attachment to secondary symbols (e.g. Savage and Gabriel, 1976). Actually, Shils and Janowitz (1948), while noting the apolitical attitudes of German soldiers, noted as well their personal devotion to national leaders. Similarly, in the case of the U.S. forces, Shils (1950) noted that the "tacit patriotism" of the American soldier contributed to combat motivation. Moskos (1970), who in fact questions the importance of primary groups, discussed the importance of "latent ideology" for American troops in Vietnam. Indeed, Moskos (1976) viewed ideology as more important than group cohesion for the combat motivation of American soldiers in World War II, Korea, and Vietnam. In the Soviet Union, the importance of ideology in the motivation of soldiers has been manifested in active programs of political socialization both in support of, and within, the military (Jones and Grupp, 1982). Indeed, this may be the major personnel issue on which Ivan stands taller than Johnny.

The depoliticization of the military was consistent with another emerging trend in American society: the advent of the welfare state. The growth of a recognition on the part of the state of its responsibility for the well-being of the citizenry has contributed to the material well-being of the populations of those societies subscribing to this ethic. This has been reflected in the allocation of governmental resources. In the United States, for example, federal government expenditures for health, social security, welfare, and labor increased from 7.7% of all federal expenditures in 1952 to 31.7% in 1975 (The Conference Board, 1977:39). The growth of the welafe state in America had two major implications for military manpower policy. First, whereas previously citizenship had been conceived as involving obligations and rights, the welfare state emphasized rights almost to the exclusion of obligations. Second, the grants economy that the welfare state spawned increasingly defined the relationship between the citizen and the state in terms of a cash nexus. The major obligations of the state were to be met by granting aid to individuals and groups of citizens, and the major obligation of the citizenry was to refill the public coffers by paying taxes. This perspective served as a strong base for the econometric blueprint of the Gates Commission that paved the way for the all-volunteer force, which left it to the labor market to decide who would serve, so that wars would be fought by those who needed the work.

The major intellectual debate on military manpower policy during the past two decades has been between those who view service to the nation, both in the military and in other roles, as a citizenship obligation (e.g. Janowitz, 1967), and those who view military manpower as a problem in labor market dynamics (e.g. Friedman, 1967). While the economic model has dominated military manpower policy during the all-volunteer force era, we have learned of its shortcomings. We have learned that in a healthy economy, it is difficult to bring quality personnel into the military in sufficient numbers to man the force, and we have learned that while taking advanatge of economic malaise allows us to upgrade the quality of the force, we do so at a cost of confronting the American people with very hard choices to make between guns and butter, because it is precisely in times of economic hardship that welfare demands on federal funding are highest (See Harries-Jenkins, 1981). Moreover, although we have been able to improve the intellectual quality of the force, the military effectiveness of the force has not been tested. And if the economy's wounds heal, we may again confront quality problems.

One solution to the problem is to keep the economy in shambles, but there are political costs to a chief executive who takes such a tack. Another strategy is to reinstitute conscription, which provides some protection from labor market fluctuations, and also yields some dollar savings. We learned during the 1960s and early 1970s, however, that the American people have little tolerance for an inequitable draft. One might argue that the inequities of the Vietnam era draft can be corrected by closer approximation to a purely random lottery, but I would suggest that the percentage of the population at risk that is actually drafted is a more important parameter than the fairness of the draw. Conscription in America has been most acceptable during those periods when it brought into the military the largest percentages of eligible young men. The worst year of the all-volunteer force saw a shortfall of about 26,000 men, and I submit that a draft of only 26,000 men, no matter how random, will not be acceptable to the American people. Randomness is not the same thing as equity.

But if the armed forces can use no more than 26,000 draftees, how can we increase the number who serve? I suggest that it be done by embedding the notion of military service in a broader matrix of national service, in which doing something for the nation becomes a normative part of American life. It may not be a cost-effective way of meeting specifically military manpower problems, but I believe it will make us a healthier nation.

REFERENCES

Bachman, J.G., Blair, J.D., Degal, D.R. 1977. The All-Volunteer Force. Ann Arbor: University of Michigan Press.

du Picq, A. 1958. Battle Studies, trans. J.N. Greely, R.C. Cotton. Harrisburg: Stackpole.

Friedman, M. 1967. Why not a volunteer army? Pp. 200-207 in S. Tax, ed., The Draft. Chicago: University of Chicago Press.

Ginzberg, E., et al. 1959. The Lost Divisions. New York: Columbia University Press.

Harries-Jenkins, G. 1981. Armed forces and the welfare state. Pp.231-257 in M. Janowitz, ed., Civil-Military Relations. Beverly Hills: Sage Publications.

Janowitz, M. 1967. The logic of national service. P.73-90 in S. Tax, ed., The Draft. Chicago: University of Chicago Press.

Janowitz, M. 1975. The all-volunteer military as a "sociopolitical problem. Social Problems 22.

Janowitz, M. 1982. Consequences of social research on the U.S. military. Armed Forces and Society 8: 507-524.

Jones, E. and Grupp, F.W. 1982. Political socialization in the Soviet military. Armed Forces and Society 8:355-387.

Marshall, S.L.A. 1947. Men Against Fire. New York: Morrow.

Moskos, C.C. 1970. The American Enlisted Man. New York: Russell Sage Foundation.

Moskos, C.C. 1976. The military. Annual Review of Sociology 2: 55-77.

Savage, P.L. and Gabriel, R.A. 1976. Cohesion and disintegration in the American army. Armed Forces and Society 2: 340-376.

Segal, D.R. 1982. Sociology in national defense. In H. Freeman et al., eds., Applied-Side Sociology. New York: Jossey-Bass.

Segal, D.R. and Segal, M.W. 1984. Changing military organization. Annual Review of Sociology 9. In press.

Shils, E.A. 1950. Primary groups in the American army. In Continuities in Social Research, ed. by R.K. Merton and P.F. Lazarsfeld. New York:Free Press.

Shils, E.A. and Janowitz, M. 1949. Cohesion and disintegration in the wehrmacht in World War II. Public Opinion Quarterly 12: 280-315.

Stouffer et al., 1949. The American Soldier, vol II, Combat and its Aftermath. Princeton: Princeton University Press.

Toomipuu, J. 1981. Soldier Capability-Army Combat Effectiveness (SCACE). Ft. Benjamin Harrison, Ind. U.S. Army Soldier Support Center.

Wallace J.R. 1982. The Gideon Criterion: The Effects of Selection Criteria on Soldier Capabilities and Battle Results. Ft. Sheridan, Il. U.S. Army Recruiting Command.

AD P000822

Military Service in the 1980s: Perspectives on the All-Volunteer Force

by

Wayne S. Sellman
Office of the Assistant Secretary of Defense
(Manpower, Reserve Affairs, and Logistics)

Transition

On 30 June 1973, Dwight Elliot Stone assumed an important, if little recognized, role in American history. He was the last U.S. citizen to be drafted. Since then, the Armed Services have been manned with volunteers. Based on the experiences of the last nine years, the Department of Defense is convinced that this is both the right way and the best way to man America's peacetime professional military.

Debate regarding what type of standing military the United States should have is not a new phenomenon. Its genesis is rooted in our colonial experience and constitution. The first major initiative to move away from a small volunteer armed force came as a result of experience in World War II and with it the realization of America's world leadership role. It is important to remember that from an historical perspective our reliance on conscription was an exception, not the rule, as a personnel procurement policy.

In March 1969, President Richard M. Nixon appointed a distinguished commission under Thomas S. Gates, former Secretary of Defense, to "develop a comprehensive plan for eliminating conscription and moving toward a volunteer force". The commission identified problems associated with the draft-dominated military manpower system that was being openly challenged by significant portions of our society, and they structured concrete proposals to be considered as alternatives to it. These proposals to meet the bonafide manpower requirements became the blueprint for a voluntary military.

What many may not know is that DoD appointed its own review force in April 1969 to "develop a program to meet future quantitative and qualitative manpower requirements to the greatest extent possible, without reliance on the draft". The DoD group provided a working dialogue with the Gates Commission, but a DoD report was not officially released to avoid the appearance of competition.

However, there was consensus in both groups as to the feasibility of the all-volunteer force (AVF) and the principal steps needed to end reliance on the draft. These included: (1) substantial pay increases for junior enlisted personnel, (2) selective pay incentives for specialists, (3) additional ROTC scholarships, (4) greatly expanded recruiting programs, (5) need to retain members of the career force, (6) preserve strength of reserve forces, and (7) special emphasis required for physician and dentist recruiting/retention programs.

On 23 April 1970, the President announced in favor of the AVF, and DoD moved from planning to action. Nevertheless, it took almost three years before AVF implementation -- a period marked by great uncertainty. During this period, there were also several decisions to promote fiscal constraint made by the Nixon Administration that undermined the expansion programs necessary to underwrite the volunteer concept. This was viewed by many manpower analysts and AVF critics as evidence that the Administration lacked real commitment to its own program.

Thus, while the transition, was uneven, by July 1973 we had moved to the AVF, and Dwight Elliot Stone was the last conscript in the United States. There was little doubt at that time that the decision reflected a broad national consensus against conscription. Even so, the great AVF debate started almost immediately.

<u>Status of the All-Volunteer Force</u>

To sustain an AVF is probably the most complex and demanding task that the Department of Defense (DoD) will face over the next decade, especially as the result of our conscious decision to increase the size of our military by approximately 200,000 servicemembers. From a policy point of view, the most important military manpower questions for the 1980s include, "Are we recruiting and retaining enough high-quality people to meet our national security requirements, and what steps must we take to ensure that we will be able to do so throughout the 1980s?"

<u>End-Strength</u> -- In every year of its existence, the AVF has either achieved the Congressionally authorized end-strength, or been no more than 1.5 percent short. It is true that during the post-Vietnam era, end-strengths were gradually reduced because of budgetary shortages, Congressional restrictions, and changes in force structure. Nevertheless, maintaining our numerical objectives so well without any resort to conscription was no small achievement. This is the only time in our nation's history that we have built a large peace-time standing force exclusively with volunteers.

<u>Recruiting</u> -- Fiscal year (FY) 1979 was the first year in which AVF recruiting did not meet planned objectives; in fact, it was seven percent short. However, because fewer people left the military that year than were expected, overall end-strength was only slightly below authorization. There is no doubt, however, that FY 1979 was the worst recruiting year in the history of the AVF.

Fortunately, the picture has brightened. In FY 1980, the Services not only met their recruiting goals but were able to make up for the previous year's shortages. This success was attributable largely to three factors: relatively high unemployment rates, particularly among youth; some recruiting innovations; and the Army's willingness to accept large numbers of high school dropouts and people who scored comparatively low on the enlistment test. As for FYs 1981 and 1982, numerical objectives were satisfied with significant improvements in the educational levels and aptitude test scores of new recruits.

This later point leads usefully away from recruit quantity to quality. The issue of quality has become one of the thorniest and most argued in the entire debate about the AVF. "Quality" is generally used by manpower analysts to describe those characteristics and attributes of military personnel that contribute to a productive, effective, and motivated force. Although many research efforts have been conducted and are underway to define and refine measures of "quality," the current operational definition of the quality of enlistees consists of two measures: educational attainment and enlistment test scores.

The Armed Services place high premium on completion of high school for the enlisted ranks. The possession of a high school diploma is the best single measure of a person's potential for adapting to life in the military. Enlistees who have not completed high school (at time of entry), are about twice as likely as are high school graduates to leave the military before finishing their first term of service. Thus, one practical gauge of military recruiting "success" has been the proportion of high school graduates.

The Military Services attempt, in any given year, to recruit as many high school graduates as possible. In some years they have been more successful than in others. Indeed, in FY 1974 only half of all Army and Marine Corps enlistees were high school graduates. However, by FY 1981 the proportion of recruits with high school diplomas had increased in all Services. In FY 1982, those percentages were the highest ever, including periods of conscription. Never before had the proportion of new recruits in the Army--or the proportion for all Services combined--eclipsed the 80-percent level. In FY 1982, those percentages were 86,79 86, 94, and 86 for the Army, Navy, Marine Corps, Air Force, and total DoD, respectively.

As in the case of formal education, the Military Services would prefer to recruit the "most trainable" young men and women from the general population. The test used to screen applicants for enlistment is the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB consists of ten subtests. The scores of four of the subtests (word knowledge, paragraph comprehension, arithmetic reasoning, and numerical operations) are combined to produce an Armed Forces Qualification Test (AFQT) score. The AFQT score, supplemented by scores on various composites of aptitude subtests, is used in conjunction with educational, medical, and moral standards to determine an applicant's enlistment eligibility. The Services prefer to enlist individuals with high AFQT scores because those recruits can be trained more quickly and are more likely to qualify for specialized training in more occupational areas.

For reporting purposes, scores on the AFQT traditionally have been grouped into five broad categories. Persons who score in Categories I and II are above average in trainability; those in Category III, average; those in Category IV, below average; and those in Category V, markedly below average and, under current Service policy, not eligible to enlist.

A recent error in calibrating the AFQT produced higher scores for many individuals than they should have been given. As a result, the Services accepted large numbers of people who would not have been

eligible to enlist had their ' ·res been calibrated properly. Particularly at the lower end of t\ scale, the error had significant consequences. Whereas we originally believed that six percent of all DoD recruits in FY 1980 were in Category IV, after correcting for the calibration error, 31 percent of all DoD recruits that year were Category IV. For the Army, 50 percent of its FY 1980 accessions scored in the below average range. The calibration problem was corrected in October 1980 with the introduction of a new ASVAB.

Increased attractiveness of military service, brought about by recent initiatives, such as the 1980 and 1981 pay packages (compensation and bonuses), innovative recruiting strategies, a test program of enhanced educational benefits, and the economy resulted in significant improvements in the AFQT scores on new recruits. The proportion of non-prior service accessions scoring in Category IV declined to 18 percent for total DoD in FY 1981 and to 13 percent in FY 1982. For the Army, the FY 1981 rate was 31 percent; that percentage dropped to 19 percent in FY 1982.

Retention -- The heart and soul of any military organization is the career force. The composition of the career force is almost completely independent of the way in which people are brought into the military for their first term. If serious problems in retaining careerists should occur, they would not be solved by a draft.

Today's area of concern is the mid-career force--those with more than 10 years of service--especially in certain critical job skills. Low first-term reenlistment rates during Vietnam coupled with declining second and third reenlistment rates since the mid-1970s have produced a force dangerously short of midcareer, senior enlisted personnel. Career reenlistment rates dropped from 80 percent in 1974 to 68 percent in 1979.

The reasons for this sharp decline are not obscure--pay scales increasingly less competitive with the private sector (in stark contrast to the explicit assumptions behind the AVF) and a general deterioration in the living conditions for military personnel and their families. Military pay kept pace with the civilian sector only for the first two years of the AVF. Pay caps in 1975, 1978, and 1979 yielded military pay in 1980 that was 20 percent below what it was in 1972. The gap between military and civilian pay had widened so much that even the substantial raises of 1980 and 1981 left military pay still behind its 1972 relationship to civilian pay. The end result has been a 'icious cycle in which mid-career shortages force those mid-career ersonnel who stay to work longer hours, serve longer overseas tours f duty, and, in the case of the Navy, have more frequent and longer ours at sea-thus discouraging many of them from reenlisting.

The 1980 and 1981 pay raises and other initiatives have tried · interrupt this cycle, and results are now beginning to be realized. reer reenlistment rates climbed to 82 percent at the end of FY 1982. t the same time, the reenlistment rates among first-termers increased om 30 percent in FY 1976 to 39 percent in FY 1980 and to 55 percent FY 1982.)

We have also paid more attention to quality of life both here and overseas, and this means, among other things more and better housing, improved medical care, and enhanced recreational facilities. But it will take a long time to repair the cumulative damage of these shortages. You do not produce a seasoned first sergeant overnight, and you cannot pick up 20,000 experienced petty officers in a year. We are moving in the right direction, but we must maintain the momentum.

Representation

Two other issues warrant consideration--women in the military and the representativeness of the force.

In 1972, women constituted 1.5 percent of the Armed Forces; today, 8.5 percent. Dramatic increases in the number of military women are the result of two developments--the women's movement throughout our society and the All-Volunteer Force. This expansion of opportunities for women in the military has been good for women, and it has been good for the military. Ethically, it is right, and pragmatically, if we are to maintain the AVF while the male youth population is shrinking, it is wise.

Our experience so far is that women exhibit the same range of competence as their male counterparts. Military women have proven themselves dedicated, effective, and professional. Yet, the ultimate issue regarding women in the military is indeed the ultimate test of a military force--combat. Thus, a comprehensive and systematic review of the role of women in the Services is underway.

The second issue is how representative the Armed Forces are of American society as a whole. The question is raised in two ways--practical and ethical.

I, for one, reject the "practical" concern based on the notion that servicemembers from certain socioeconomic backgrounds or of some races or from particular regions of the country will be less willing or able than their comrades in arms to defend America or American interests under certain war scenarios. This argument is specious at best, bigoted at worst. Based on experience in past wars and based on what I know firsthand of those in uniform today, I personally see no grounds for concern along these lines.

The ethical concern is, in theory, more well-founded. The burden of defending an entire society should not fall disproportionately on any one group or segment of that society. I say that knowing full well that virtually no army in history has been fully representative of the society it defends.

Numerous surveys and studies of the representativeness of the force have been conducted. The truth belies the popular myth. In terms of socioeconomic status, the very highest and the very lowest brackets are underrepresented in the enlisted force, but otherwise it is quite representative. Geographically, we are getting a proportionate share of recruits from all regions and all states. Our most recent major study compared 18-23 year-old military personnel with

their contemporaries in the civilian workforce. The findings will be surprising to many: (1) the percentage of high school graduates is greater, (2) the educational and occupational distributions of their parents are virtually the same, (3) their marital status distribution is the same, (4) their health profiles reveal no differences, and (5) their mental abilities are somewhat higher.

In terms of race, the minority composition of the Armed Forces began to grow during the Vietnam War, and it has increased more rapidly under the AVF. It is important to note two facts. First, since 1973, all recruits were volunteers, not draftees; and, second, higher percentages of black youth meet the standards for enlistment now than before. Improved educational opportunities for blacks have, I believe, yielded higher aptitude scores for blacks. During this same period, however, unemployment rates for black youth have become high. In my opinion, the military offers blacks and other minorities better opportunities for training and advancement than does much of the civilian sector. It is no surprise, therefore, that large numbers of blacks are joining and making a career of the Services.

At the same time, the equity issue persists--no group should have to bear a disproportionate share of the burden of defending, or, in the event of war, a disproportionate share of the casualties. I do not believe we are at the former stage yet, nor do I foresee it in the future. As for the latter, a major war would in all likelihood stimulate a draft, and racial balance among military personnel, including casualties, would be quickly restored.

Future Forecast For The AVF

I have spoken about the past and present of the AVF. In that regard, we need to recognize both its successes and problems. A la Mark Twain, the rumors of the death (or even the terminal illness) of the AVF are premature. Thus, we in DoD are convinced that the AVF is a success; however, we cannot become complacent. The recent military pay raises were essential. Educational incentives must be enhanced. Quality of life must be improved and maintained. The Reserves, in particular, must be strengthened.

Last year, the President appointed a Military Manpower Task Force, chaired by Secretary Caspar Weinberger. The Task Force has worked hard reviewing the adequacy of military compensation and incentives; educational benefits; current manpower readiness; effectiveness of training, leadership, and discipline; enlistment standards; recruiting and retention efforts; and Selective Service registration. Its findings were released in a press conference on 18 October 1982, and the report will be available shortly.

Finally, another key element important to the success of the volunteer force is the attitude of the public toward our servicemembers. Over the past several years, the American people has become more supportive of our young men and women in uniform. That positive shift in attitude, if sustained, combined with management initiatives and appropriate compensation levels should preserve the viability of the AVF.

JOB-TASK ANALYSIS: MATCHING INPUT TO OUTPUT

Chair: F. Worth Scanland

Several papers were presented. One compared costs for storing, analyzing, and maintaining a totally automated data base (to support Naval ISD) to the costs of current data analysis which includes both automated data storage and analysis and use of subject-matter experts. Another paper compared and contrasted occupational data input methodologies for both current and projected use, based on their suitability for front-end job/task/skill analysis. The third paper reflected on the current method of data collection in occupational analysis and compared it to other data collection methodologies for cost, objectivity, and kind of information obtainable.

FRONT-END ANALYSIS: CONFIGURING OCCUPATIONAL DATA INPUT TO SUIT OUTPUT

Thomas M. Ansbro
Headquarters Chief of Naval Education and Training
Naval Air Station Pensacola, Florida

This paper discusses Navy occupational data input methodologies in both current and projected use, related to their determined suitability for employment in front-end job/task/skill analysis (FEA). The text is essentially a historical treatment of a practicum based on ideas presented in the paper titled "Selection of a Data Collection Methodology for Occupational Analysis" by William A. Hayes.

Projected data input and analysis methodologies are discussed in view of potential support or produces outputs projected downstream. These are seen to exist at two levels: foreseen outputs of the entire front-end analysis system (job/task inventories for paygrades, billet descriptions/manpower documents, training programs, curricula) and those within the analysis subsystem (of the processing mechanism itself--task/skill hierarchies, inter/intra task relationships, rankings, etc.).

FEA PARAMETERS

In 1975 a Navy training systems research organization tasked to consolidate Navy Occupational Standards and other Navy occupational data for use as a "front end" to the then-new Instructional Systems Development (ISD) System addressed its assignment by setting parameters for an "ideal" FEA:

a. The Navy enlisted "world of work" must be the source of all procured occupational data.

b. All data should come from officially documented sources or existing "hard data" (work-record) systems.

c. Occupational data input should be essentially "raw", devoid of any processing of evaluative or judgemental criteria. Judgemental data should be products of analysis, preferably by a process that could be procedurally mechanized.

d. The FEA should provide a Navy occupational data base suitable for use beyond application of ISD. The rationale is that training programs derive from world-of-work data principally aligned with job/billet requirements and job-incumbent certification. In order, job requirements influence (if not clearly dictate) the skills mastery array of the job incumbent; certification and advancement align with this mastery, and training programs are designed specifically to provide this mastery. Therefore, an occupational data base and analysis system could not be designed to meet the needs of ISD alone without also attending to basic job-descriptive analytic needs of manpower management and job-incumbent certification/advancement, as well as training.

e. The training community requires FEA outputs in greater detail and depth then other users (d above) of occupational data; therefore, meeting the data and analysis needs of trainers should satisfy other user requirements as well.

At this point in the development of further parameters, clarification was needed concerning what constituted descriptive occupational data and what was the product

of job/task analysis (Rankin, 1975). Most work (task) data extant appeared to be a mixture of both. Questionnaires distributed to the fleet gleaned an admixture of job-incumbent responses that covered frequency of task performance, percentage of time spent performing, equipment worked with or on, tools, instruments used (recall of basically factual information), job importance/satisfaction (opinion), physical/mental characteristics (experience, judgement), and task lists augmented by supporting skills. These data essentially provided inventories of tasks, skills, and underlying work-performance characteristics. Follow-on analyses statistically arranged, rearranged, and prioritized elements in these inventories for the indicated data consumers throughout the Navy. Certainly, almost everything that trainers needed appeared to be there, even references; but a hierarchical information structure keyed to individual tasks in an inventory was not. By whatever means the data were acquired, the line between task description (input?) and analysis (output?) was not clear.

WORK-UP OF FEA MECHANISM

In an attempt to establish separate classifications for occupational data input and analysis, it was determined that all task-descriptive data in a job/task inventory (JTI) would be provided (acquired) as input--no data requiring conjecture, judgement, evaluation, comparison, ranking, or assignment to paygrade or skill level--only data coming directly from official Navy technical documentation and Subject-Matter-Expert (SME) recall of documented on-job experience (tasks, task elements, component skills, task action objects, references, tools, equipment, support materials). In the resulting JTI mockup, individual-task entry format (Figure 1) was composed of separate "data blocks":

1. Categorical: The task statement itself (action and object of action).
2. Environmental: In what work-site category or environment (platform, system, equipment, shop, etc.) is the task performed?
3. Identifying/supporting: According to what authority/reference; with what supporting items, materials, tools, equipment is task performed, and according to what standard?
4. Descriptive: What detailed component, subordinate, or supporting work behaviors describe or underlie performance of the task (thereby providing skills profile of a job incumbent)?

These data blocks were keyed to "specific-action/object" task statements rather than the "generic" statements (frequently specific action, but "typical" or "representative" object of action: electric motors, generators, ground support equipment, etc.) common to then-current occupational data banks. Consequently, the proposed data acquisition presupposed considerably larger JTIs and repetitive task-descriptive input. However, the data input was scheduled for use with tabulated data-entry forms and was to be essentially a data transfer from technical documentation, augmented by SME job experience. The anticipated volume of tasks to be listed and task-descriptive data to be recorded ruled out any prospective use of questionnaires or observation of incumbents at work. Scope of any JTI would most likely be bounded by such Navy entities as rating or Naval Enlisted Classification (NEC). Assembling JTIs would eventually comprise an occupational data base.

It was decided that the data base provided by such an assembly should be complemented by a FEA system capable of operating within, among, and across occupational fields and yielding output data of sufficient specificity to support

writing billet descriptions, determining job-incumbent task/skill performance requirements, and developing job/skill training programs. Further, data acquisition, storage, and analysis should provide an independent "front end" for ISD rather than be incorporated into it. ISD was seen to use processed as well as descriptive data: task distribution, hierarchies, interrelationships, priorities, rearrangements, etc; therefore, the relationship of FEA to ISD could be shown as "producer-to-consumer". If the descriptive data input were sufficiently extensive and detailed, with component features catalogued, most (possibly all) judgemental outputs (hierarchies, etc.) should be produced by computer, thereby reinforcing the producer-to-consumer concept and literally taking the instructional system/curriculum developer (ISD user) out of FEA as a producer.

OUTPUTS OF ANALYSIS

Outputs of FEA should address such foreseen interrelationships among data elements/items as commonality and componency; further, the perceived complexity inherent in each task should be calculcated and recorded:

1. COMPLEXITY: Numerical index determined by quantifying task-descriptive data. Task complexity should be a fixed factor, dependent upon measurable physical and mental characteristics of task performance requirements and component skills (Figures 1 and 2). It should not be described as "learning difficulty" or "task difficulty"; such factors appear to be variables influenced by characteristics of the task performer as well as by the inherent complexity of the task (Ansbro, 1977). Complexity determination should be made by the computer early in the task-analytic process, since it would become a structural entity in producing other related outputs (Figure 1). Complexity is seen as an element in a vertical hierarchy of task ranking.

2. COMMONALITY: Task-to-task relationship, determined by matching identifying and descriptive factors task-by-task. Whether or not a categorical task statement (action + object) matches that of another task (or of many others), the component descriptors (conditions, standards, underlying skills) should be the factors upon which a commonality decision depends. Tasks should be considered common if all the descriptors of one identically match all those of another. Commonality could be further delineated by some producer-consumer agreement on a degree or percentage of task similarity somewhat below the stringent requirement of "identical". It should follow that identical tasks have identical complexity indices; if so, then one computer output would appear to verify the other. Commonality is seen as lateral distribution throughout occupations. A detailed examination of task-descriptive data would likely disclose a high degree of task commonality within and across ratings/NECs. Such findings would help to compress or reduce the size of the data fields, facilitating FEA data-handling and management in spite of the projected size and coverage of JTI data acquisition.

3. COMPONENCY: A vertical hierarchy of work-behavior coverage, or span, within which tasks of greater span (and attendant higher complexity indices) superimpose on those of lesser span, provided that all work behaviors of those tasks of lesser span are included in those tasks of greater span. It is a proposition that in a hierarchy of tasks so arranged, large-scope, high-complexity, multi-behavior-encompassing tasks would contain (or "embody") tasks of lesser content, permitting computer scanning of task content to group tasks in such content hierarchies. Resultant outputs of such analysis should become a salient feature in prioritizing tasks for training or determining skill acquisition spectra for enlisted careers. With regard to one principal benefit to be

derived from the computer's projected exercise of commonality, the employment of componency hierarchies should further compress the size of processed data inventories to be used by FEA consumers.

4. CRITICALITY: Some measure of the importance of performing a particular task to the completion of a larger function (job, assignment, etc.). To what degree would inadequate performance of a task decay the quality of job performance? Or impact on safety? Could such impact be accurately measured (especially incrementally)? Criticality appears to be a factor in task element and skill performance as well.

ADD-ONS

A fifth data block category was added to the JTI input document--Adjunctive: What further information (not descriptive of supporting work behaviors) helps to detail or further define task-performance environment? Block content reflected data collected in other FEA systems as well as some internal outputs of the proposed FEA (Figure 1).

TRIAL RUNS OF MECHANISM

With the basic design of the FEA system completed and all computer-programming requirements of the mechanism developed, FEA was given a series of trail runs. Salient results follow:

a. Internal outputs (complexity, commonality, componency) proved workable and useful. Their employment made possible computer separation of job-specific tasks and rating-specific skills, disclosed high incidence of task commonality within and across assumptively associated ratings, made possible task distribution among paygrades, provided prioritized task/skill lists for training purposes.

b. Criticality, as an internal output, was never successfully employed, was discontinued. Flexibility of analysis and the success of other outputs accelerated departure of the rest of the Adjunctive data block.

c. It proved possible for the computer to translate task statements (with included component behavior descriptors) into learning objective format, thus, in effect, printing-out curriculum outlines. The computer also printed out billet descriptions (in terms of tasks to be performed by incumbents).

d. Occupational data collection by questionnaire for the subject FEA was determined to be impractical. Input data structure required listing tasks with specific action objects and citing all appropriate descriptors provided. The resulting volume of data input (Figures 1 and 2) (23,000 tasks in 4 Navy ratings) dictated a less manpower-intensive method. However, man-hours expended in the selected method did not exceed that total expended by then-current data collection methodology.

(1) Resulting JTIs for individual ratings ran to several thousand tasks listed, reasonable coverage of each rating.

(2) Tabulated-data entry forms for manual service (input) eventually gave way to use of a microcomputer for input, facilitating and speeding up the input process.

REFERENCES

ANSBRO, T.M.: Commonality and complexity as Factors in Skills Transferability, Proceedings - 19th Annual Conference of the Military Testing Association, Oklahoma City, OK, 1978 263-274.

RANKIN, W.C.: Task Description and Analysis for Training System Design, TAEG Technical Mamorandum 74-2, 1975

```
TRAINING DEVELOPMENT          TASK INVENTORY FILE          DATE:  760804

TASK REFERENCE NO.   TASK ACTION CODE   DUTY SUBCATEGORIES   PAYGRADE   SKILL LEVEL   OCSTD NO.
AE 0002 0034         REP                GENERAL, 02          E-5        JOURNEYMAN    AE24332

ACTION:  REPAIR PWR LVR HORN & SW ASSY  1   3-M WUC 29319

PLATFORM.....P-3 A/B/C AIRCRAFT              SYSTEM.....T-56 ENGINE
EQUIPMENT....THROTTLE/PWR LEVER              COMPONENT..PWR LVR WRN HORN ACTR & SW  2
MODULE.......NC

CUES..................MALFUNCTION, REPAIR COMPLETE
REFERENCES........NA 01-75PAA-2-4
STANDARDS.........IAW REFERENCE PU38
TOOLS.............COMMON HAND TOOLS
SUPPORT MATLS.....CROCUS CLOTH, CLN SOLV P-DP680, LUBOIL VVL-800  3
SUPPORT EQUIPT....COMPRESSED AIR/NITROGEN
TEST EQUIPT.......MULTIMETER AN/PSM-4

GENERAL...........21131111120000000000       COMPLEXITY INDEX     1.25
DUTY SUB 02.......02000012000000000000       COMMON TO            205(WIR) 1706(ASR)
SKILL 1...........21113000000000000000       COMPONENCY           OMNI
SKILL 2...........12110000000000000000       CRITICALITY          05
SKILL 3...........20000000000000000000       FREQ/PERF            03
SKILL 4...........11100110000000000000       MAN-HR REQ           01.5 HR(SW)
SKILL 5...........22332111300000000000  4    CONSEQ INAD PERF     03              5
                                             TRAINING INDEX       04P
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

LEGEND: 1 CATEGORICAL 2 ENVIRONMENTAL 3 IDENTIFYING/SUPPORTING
4 DESCRIPTIVE 5 ADJUNCTIVE

FIGURE 1. Computer Printout of Individual Task Listing in FEA JTI

TASK-DESCRIPTIVE CHARACTERISTICS (DESCRIPTIVE FACTORS)

MFC (O) ALL
DSC (X) GENERAL

THESE FACTORS DEAL WITH ALL ASPECTS OF A TASK, THE OBJECT BEING WORKED ON, SUPPORT EQUIPMENT, TEST EQUIPMENT, TOOLS, ETC. FOR ALL TASKS, MARK THE FACTORS IDENTIFIED BY A DOUBLE ASTERISK (**) PLUS WHATEVER ADDITIONAL FACTORS PERTAIN.

A. FOLLOW SAFETY PRECAUTIONS (SAFETY, CONSEQUENCES) ** (CONCERN DIRECT ACTIONS IN PERFORMING THE TASK)

(2) VIOLATION MAY CAUSE MINOR DAMAGE TO EQUIPMENT AND/OR MINOR INJURY TO PERSONNEL.

B. RECOGNIZE/OBSERVE SAFETY PRECAUTIONS **

(1) RECOGNIZE ESTABLISHED SYMBOLS,. SIGNS, LABELS; OBSERVE ESTABLISHED PRECAUTIONS.

C. RESPOND TO INITIATING CUES (SPEED AND CONTINUITY) **

(1) IMMEDIATE RESPONSE IS NOT REQUIRED, AND ACTION MAY START AND STOP WITHOUT AFFECTING OVERALL PERFORMANCE (ALLOWS TIME FOR RESEARCH BEFORE AND DURING TASK PERFORMANCE).

D. GAIN ACCESS TO TASK OBJECT (ACCESSIBILITY) **

(3) DIFFICULT TO GAIN ACCESS, E. G., REQUIRES DISASSEMBLY OR REMOVAL OF OTHER COMPONENTS.

E. MAKE ADJUSTMENTS OR SETTINGS/TAKE MEASUREMENTS (TOLERANCE OR PRECISION WITH WHICH SETTINGS ETC. MUST BE MADE)

(1) APPROXIMATE/COARSE----NO ESTABLISHED TOLERANCES.

F. VIEW/MANIPULATE ITEMS (MINIATURIZATION)

(1) SIZE IS OF LITTLE CONSEQUENCE IN POSITIONING.

G. HANDLE/REPOSITION ITEMS (SIZE/SHAPE)

(1) SIZE/SHAPE IS OF LITTLE CONSEQUENCE IN POSITIONING.

SKILL AREA: (5) USE TEST EQUIPMENT

NOTE: FOR ALL TASKS, MARK THOSE FACTORS IDENTIFIED BY A DOUBLE ASTERISK (**).

A. SELECT TEST EQUIPMENT **

(2) REQUIRES A SINGLE ITEM OF TEST EQUIPMENT WHICH SERVES MORE THAN ONE PURPOSE, E. G. MULTIMETER, TEST SET, ENGINE ANALYZER, OR,

REQUIRES MORE THAN ONE ITEM OF TEST EQUIPMENT---EACH SERVES ONLY ONE PURPOSE, E. G., A FEELER GAUGE AND A DWELL METER, A TORQUE WRENCH AND A MICROMETER, A SIGNAL GENERATOR AND A WATTMETER.

B. POSITION TEST EQUIPMENT (MOBILITY AND WEIGHT) **

(2) TEST EQUIPMENT IS MOBILE AND WEIGHT ENABLES ONE PERSON TO REPOSITION, E. G. MULTIMETER, AIR PRESSURE GAUGE.

C. OPERATE TEST EQUIPMENT (CHARACTERISTICS)

(3) MUST BE CONNECTED TO SYSTEM AND REQUIRES MANUAL STEP-BY-STEP PROCEDURES TO OBTAIN READINGS.

NOTE: ITEMS D, E AND F REFER TO TEST EQUIPMENT, NOT TO EQUIPMENT BEING WORKED ON OR OPERATED, NOR TO SUPPORT EQUIPMENT.

D. MANIPULATE CONTROLS (NUMBER AND TYPE)

(3) MULTIPLE CONTROLS---MULTIPLE TYPES, E. G. COMBINATION OF SWITCHES, KNOBS, LEVERS.

E. OBSERVE/ANALYZE INDICATORS (NUMBER AND TYPE)

(2) MULTIPLE INDICATORS---SINGLE TYPE, E. G. MULTIPLE DIALS, LIGHTS, GAUGES.

F. OBSERVE INDICATORS (LOCATION)

(1) VIEW ALL INDICATORS WITHOUT PHYSICALLY CHANGING POSITION.

G. OBSERVE INDICATORS (READOUT)

(1) STATIONARY READOUT AND REQUIRES NO INTERPOLATION BETWEEN GRADUATIONS/SCALES, E. G. LIGHT, DIGITAL READOUT.

\- -

FIGURE 2: Narrative-Format Computer Printout of Task-Descriptive Data (checked data blocks in Figure 1)

TOTALLY AUTOMATED VS CURRENT DATA ANALYSIS FOR NAVY ISD: A COST COMPARISON

Douglass Davis, Ph.D.
Head, Research Applications
Chief of Naval Education and Training
Pensacola, Florida

ABSTRACT

This paper compares costs for storing, analyzing, and maintaining a totally automated data base to the costs incurred in utilizing current data analysis that includes both automated data storage and analysis and use of subject matter experts. The Naval Education and Training Command has used documented maintenance and repair procedures from four avionics (aviation electronics) ratings to build a computerized data base for automated job/task/skill analysis in the ISD process; current analysis uses data collected primarily from job incumbents, stored and analyzed in a computer, and analyzed further by users for employment in ISD. Costs are compared over a ten-year economic life for the Aviation Electrician's Mate (AE) Rating.

The third in a trilogy of Navy papers addressing training job/task/skill analysis methodology, this paper compares the costs for storing, analyzing, and maintaining a totally automated data base for the Aviation Electrician's Mate (AE) Rating to the costs incurred in utilizing current data analysis that includes both automated data storage and analysis and use of technical subject matter experts (SMEs).

The present method of job/task/skill analysis in Navy Instructional Systems Development (ISD) makes extensive use of data supplied by the Navy Occupational Task Analysis Program, (NOTAP). The NOTAP collects and processes data about job tasks reported to be performed by job incumbents. These data are obtained periodically through surveys, stored in a computer, and made

available to the Naval Education and Training Command in a standard package. NOTAP printouts are utilized in combination with equipment technical manuals and experienced petty officer (SME) judgment. Additional information is often extracted from such sources as the Ships Equipment Configuration System, Weapon System File, and Fleet Modernization Program. Data from all sources are reviewed and the SMEs select from among the data tasks recommended for training in Navy courses.

The Naval Education and Training Command experimental Task Inventory File was designed to make job/task/skill analyses through use of data taken exclusively from official equipment manuals and other approved sources of job data. This method utilizes SMEs to record job data, by equipments, onto Job Data Worksheets, which include descriptive information: what the task is, where it is performed, the cues which initiate performance, and other requirements, including tools, equipments, materials, etc. The SMEs also record the duty subcategory for each task, which is a further breakdown of a major functional category, such as "maintenance." For each task recorded on a Job Data Worksheet, a technician completes a Task Data Worksheet containing up to one-hundred-fifty-six descriptive characteristics that define the task. These descriptive data are scanned by the computer which orders tasks by componency, commonality, and complexity indicator. Directly from these printouts tasks are selected for assignment in Navy training courses.

For purposes of comparing costs attendant the job/task/skill analysis method presently in force and the costs attendant to the totally computerized method, each method was treated as an alternative. The present method is referred to as Alternative I; the computerized method, Alternative II. These two alternatives were the only two considered since they were the only known methods of job/task/skill analysis in Navy ISD (Fink, 1978).

Approach

The approach used in this cost analysis was that of examining actual costs of the two methods for job/task/skill analysis for the AE Rating. The resource requirements for each alternative were identified, and those common for both alternatives were factored out of the analysis. Research and development costs which had already occurred were treated as sunk costs and were considered to be irrelevant in the cost analysis. Those resources required by, but not common to, each

alternative were identified, quantified, and costed, and the costs for each alternative were specifically identified.

Assumptions

The analysis of alternatives was subject to the following assumptions:

1. The facilities required for each of the two alternatives will be approximately the same; facility costs will not be included as a data element in the analysis.

2. Both alternatives would continue to function as presently configured: combined in-house assets, primarily personnel, and data services obtained through contract or other agreement; therefore, investment costs will not be included in the comparisons.

3. The costs for future years are to be discounted at an average rate of 10 percent.

4. The costs shown reflect only the appropriate differences between the two methods and are not total costs.

Costs

Alternative I: Continue the Present Method. The projected costs for Alternative I are presented in two categories, personnel related costs, and operating costs, which include printing costs and computer services costs. Total discounted (present-value) costs for Alternative I are shown in Table 1.

Table 1

Total Discounted Costs* for Alternative I, Project Years One Through Ten ($000)

Project Year	Military Personnel Costs	Operating Costs	Cumulative Annual Costs
1	$53.74	$ 4.22	$ 57.96
2	-----	-----	57.96
3	-----	-----	57.96
4	-----	-----	57.96
5	36.73	2.88	97.57
6	-----	-----	97.57
7	-----	-----	97.57
8	-----	-----	97.57
9	25.06	1.08	123.71
10	-----	-----	$123.71

* Data sources:

Navy Occupational Data Analysis Center Memoranda, CDR TODARO, of 24 Feb and 10 Mar 1981

Navy Occupational Task Analysis Program Computer Printout AE RATING-Responses . . . by Paygrade, Skill Levels, and Total, No. 2AE-01A9

Navy Comptroller Manual, NAVCOMPNOTE 7041 of 30 Nov 1979

Personal Conversation with CNTECHTRA (Code 0162) of 5 Dec 1980

Alternative II: Implement the Computerized Model. The projected costs for Alternative II are presented in two categories, personnel costs and operating costs. All personnel costs for Alternative II are Training Commands Costs. Operating costs are printing costs and computer services costs. Total discounted (present-value) costs for Alternative II are shown in Table 2.

Table 2

Total Discounted Costs* for Alternative II, Project Years One Through Ten ($000)

Project Year	Military Personnel Costs	Operating Costs	Cumulative Annual Costs
1	$49.90	$ 6.09	$ 53.99
2	1.82	.48	56.29
3	1.66	.43	58.38
4	1.51	.39	60.28
5	1.37	.97	62.57
6	1.24	.33	64.14
7	1.13	.30	65.57
8	1.03	.27	66.87
9	.94	.63	68.44
10	.85	.22	69.51

* Data sources:
Personal Conversation with CNET (Code N-54 of 10 Mar 1981
Navy Comptroller Manual, NAVCOMPNOTE 7041 of 30 Nov 1979

Comparison of Costs

Alternatives I and II were compared in terms of discounted costs since the period of comparison was greater than three years. The cumulative annual costs shown in Tables 1 and 2 are present-value-costs for the two alternatives. That comparison indicates that the discounted cumulative annual costs for Alternative II are less than sixty percent of the discounted cumulative annual costs for Alternative I.

To show the amount of money, which, if budgeted in equal yearly installments, would pay for the alternatives, a uniform annual cost for each alternative was calculated (Naval Automated Data Command, 1980). The uniform annual cost incorporates the concept of time

value of money, whereas a simple arithmetic average of the cumulative annual costs does not. The uniform annual cost for Alternative I was $19,188.00. The uniform annual cost for Alternative II was $10,798.00.

A graphic illustration of the discounted cumulative annual costs for both alternatives is presented in Figure 1. The intersection of the cost curves in Figure 1 determines the break even point, or the point at which the economic desirability of the two alternatives is equal. That point lies between years three and four.

Selection of Alternative

The variance between the total discounted costs for Alternatives I and II over an economic life of ten years resulted in Alternative II being the more economically feasible method for collecting, processing, and retreiving data for Navy ISD job/task/skill analysis.

REFERENCES

Fink, C. D. An Analysis of CTAD instructional program development training task analysis procedures, NAVEDTRA 106A, supplement 2. Alexandria, VA: Kinton, Inc., 1978.

Naval Data Automation Command. Economic analysis procedures for automated data processing. (NAVDAC PUB 15 7000), Washington, DC, .980.

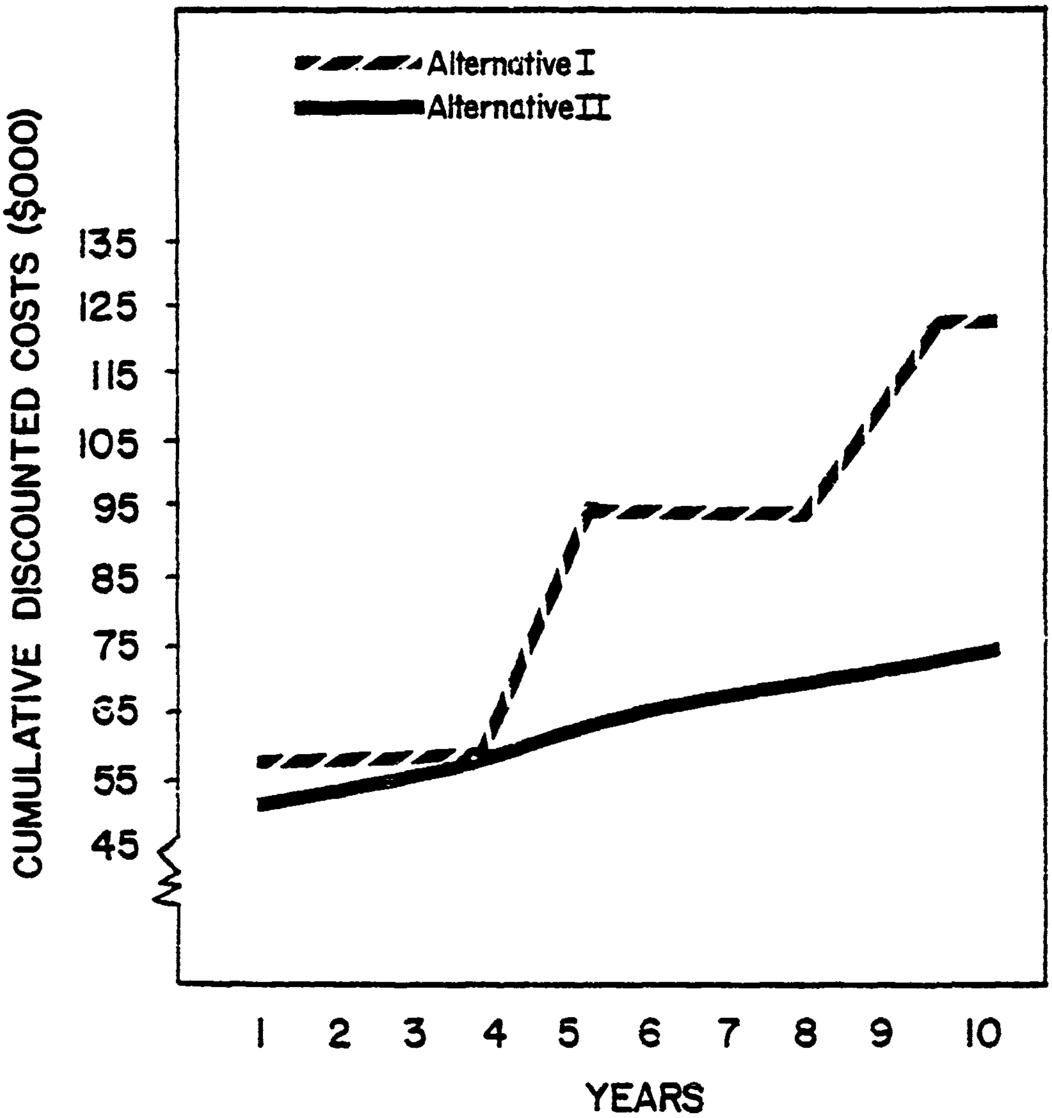


Figure 1. Break-even chart for Alternatives I and II, based upon cumulative discounted cost.

AD P000825

SELECTION OF A DATA COLLECTION METHODOLOGY FOR OCCUPATIONAL ANALYSIS

William A. Hayes
Headquarters Chief Naval Education and Training
Naval Air Station Pensacola, Florida

Monumental effort has been expended on the analysis of occupational data to produce numerous products for Manpower, Personnel, and Training Managers. In comparison, however, little has been done to analyze input methodologies and determine the optimum method for collection of that raw data. This paper reflects on the method of data collection currently in use and compares it with an alternative data-collection methodology for:

1. Kind of information obtainable,
2. Objectivity, and
3. Cost.

In order for the products of analysis to continue providing managers with adequate and timely decision-making information, the input must contain not only sufficient detail to produce those products, but it must also be objective and cost-effective as well. As the outputs become more numerous and sophisticated and their accompanying processing mechanisms are refined, the input methodology inevitably is impacted by these events. It must be reviewed and re-examined periodically to avoid being overcome by the advancing sophistication of these mechanisms.

DATA COLLECTION TECHNIQUES

Data collection techniques are widely varied and, whenever possible, tailored to the output requirements. The traditional and current method of occupational data collection is the questionnaire. It will be compared with an alternative method, the observation. In actual practice, the observation is seldom used alone and is usually supplemented by the interview. The use of the term in this paper implies the combination of both observation and interview. This combination method may be employed either in the manual mode using a subject matter expert as the investigator or in an automated mode using the computer as a partial substitute for the investigator.

KIND OF INFORMATION OBTAINABLE

THE QUESTIONNAIRE

The nature of the questionnaire dictates that it be created prior to distribution. Consequently, the originator

or creator must have a working knowledge of the subject in order to generate appropriate items upon which respondents will be allowed to vote. This technique is ideally suited for collection of some data but unsuited for other. Both the pros and the cons will be addressed.

It is extremely difficult to produce a succinct instrument, for in order to produce one which is compact and concise, it becomes necessary to compromise detail for brevity. According to Rankin (3), construction of an adequate questionnaire is seldom achieved in one iteration.

There is little personal contact with respondents during the administration of the questionnaire. Possibly due to lack of attention, respondents often misinterpret items being surveyed. Additionally according to Van Dalen (5), respondents frequently tailor replies to conform to their biases, to protect their self interests, to place themselves in a more favorable light, to please the researcher, or to conform to accepted social patterns. Such an instrument is not well suited for collection of factual "hard data" which is a matter of record and available from other sources.

These disadvantages tend to be partially offset by the major advantage of the questionnaire — information may be collected from a large number of respondents in a relatively short time. This feature makes it an ideal instrument for collection of data relating to opinions, self-perceptions, subjective judgements, attitudes, and the like.

THE OBSERVATION

Like the questionnaire, the observation technique requires that the investigator (or creator) have a working knowledge of the subject. An additional requirement, unlike the questionnaire, is that it requires the investigator to participate in a "one-on-one" situation with the respondents. This feature substantially increases the manpower expended in data collection. This disadvantage, particularly if data are required from a large number of respondents, renders it unsuited for collection of opinions, self-perceptions, subjective judgements, attitudes, and the like.

Even though many participants may not contribute during the observation process, complete and accurate data may still be collected. The investigator has the liberty of evaluating the data and separating the essential facts from the non-essential. In a face-to-face meeting, an investigator is able to probe more deeply into a problem, particularly an emotionally laden one. Questions are easily resolved and fewer mistakes or misinterpretations are encountered. Because of these factors, the data are usually

more precise and less opinionated. The observation technique is particularly well suited for collection of factual or "hard data" relating to performance, the end products of performance, or records.

OBJECTIVITY

When comparing the two techniques for objectivity, TenBrink found that the questionnaire was highly subject to bias and error, and it was less objective than the observation. The observation, however can also be subjective, especially if the instrument has been poorly constructed; but usually it tends to be the more objective collection technique.

COST

Two primary ingredients affect the cost of data collection: time and the number of personnel involved. Davis (1) found the cost of the two techniques to be approximately equal; the anomaly occurred for different reasons, however. The cost of the questionnaire derived from a large number of personnel utilized for a relatively short time. The cost of the observation derived from a few personnel utilized for a relatively lengthy time.

Figure 1 is a summary of the characteristics of the two data collection techniques.

	QUESTIONNAIRE	OBSERVATION
KIND OF INFORMATION OBTAINABLE	Self-perceptions, subjective judgements, attitudes	Performance, end products, facts, "hard data"
OBJECTIVITY	Least objective, highly subject to bias and error	Can be subjective, usually objective
COST	Many personnel, short time	Few personnel, long time

Figure 1 - Comparison of Data Collection Techniques

SPECIFIC REQUIREMENTS FOR OCCUPATIONAL ANALYSIS

Occupational analyses support the needs of manpower, personnel, and training managers. The kind of judgments and decisions made by these managers dictate the type of information, and consequently, the kind of data they require

to perform their roles. A major function of manpower managers is to determine manpower billet requirements; personnel managers assign personnel to fill those billets. A data base which supports those types of requirements has traditionally been one of broad description and whole-job detail.

A major function of training managers, on the other hand, is to provide training which will enable graduates to perform specific tasks in those billets. Driskill (2) and Rankin (3) agree that these managers need detailed descriptions of work performed by billet incumbents in order to provide the proper type and quantity of training. Too little detail in the trainer's data base can lead to incomplete training or else to costly overtraining or, the trainer simply has to go out and get the data himself.

Figure 2 depicts the scope of the data base requirements. While it might appear that the data requirements are so diverse that two data bases might be required to ensure satisfaction of all concerned, resource constraints prohibit that luxury. Is it possible then that a single data base might serve the needs of all three managers?

		BREADTH OF DESCRIPTION	
		WHOLE JOB	PARTS
LEVEL OF DESCRIPTION	BROAD	MANPOWER/ PERSONNEL MANAGERS	
	DETAILED		TRAINING MANAGERS

FIGURE 2 - Scope of Data Base Requirements

For a single occupational data base to adequately serve all manpower, personnel and training managers, it must be sufficiently detailed to provide specific needs for the most discerning or for those data users with need for greatest depth of subordinate data and detail. From specific detailed data, generalities may be derived; the obverse, however, is not necessarily true. It therefore follows that the detailed data base which would adequately support training managers could also be used to support manpower and personnel managers.

In order for that single occupational data base to be effective, it must possess the following characteristics:

1. it must describe the tasks in detail,
2. it must be derived objectively, and
3. it must be produced inexpensively.

CONCLUSIONS

Rankin (3) observed that the primary difference between the questionnaire and the observation method is that the observation method culminates in a task description, whereas the questionnaire requires one, a priori. In principle, all tasks are described and cast in questionnaire format for check-off or endorsement; incumbents enter the picture merely for obtaining estimates of the proportion of a population who actually perform or are associated with the prescribed tasks. The attitudes, opinions and perceptions available through this technique could adequately satisfy the needs of manpower and personnel managers, but not those of training managers.

The observation technique with its objectively-derived performance and "hard data" seems to be better suited for the training manager's needs. Since the cost of the two methods are approximately equal, and since the precise, detailed data provided by the observation technique can support the needs of training managers as well as those of the manpower and personnel managers, it seems obvious that it should be the method selected for support of the single data base of all occupational analysis.

REFERENCES

1. Davis, D. D.: A Comparative Study of Two Navy Instructional Systems Development (ISD) Procedures for Determining Task Commonality in the Aviation Electrician's Mate (AE) Rating. Dissertation submitted to Florida State University, 1981

2. Driskill, W. E., Gentner, F. C.: Four Fundamental Criteria for Describing the Tasks of an Occupational Specialty. Proceedings of Military Testing Association Convention, 1978

3. Rankin, W. C.: Task Description and Analysis for Training System Design. TAEG Technical Memorandum 74-2, 1975

4. TenBrink, T. D.: Evaluation A Practical Guide for Teachers, McGraw-Hill, 1974

5. Van Dalen, D. B.: Understanding Educational Research: An Introduction, McGraw-Hill, 1979

NEW DIRECTIONS IN TURNOVER RESEARCH: PROBLEMS & PROMISE

Chair: Thomas Watson

Interest in turnover is high in both military and civilian organizations. As a result of this considerable interest, turnover research has undergone a conceptual and methodological evolution over the past several years. In this symposium, the problems and promise associated with new directions in turnover research were discussed from the perspective of investigators in the Air Force and academia. A representative from the Air Force Manpower and Personnel Center (AFMPC) described new approaches to the study of turnover and retention used at the Retention Studies and Reports Division. Also, representatives from the Air Force Human Resources Laboratory (AFHRL) discussed efforts to develop a precise taxonomy of turnover criteria, to examine turnover from a dynamic process perspective, and to validate an Air Force vocational interest inventory using attrition criteria. In addition, representatives from the University of Texas at Austin discussed problems associated with the unique contribution of general and specific satisfaction to turnover decisions, and the advantages of using a "Butterfly" Catastrophe Model in turnover research.

A CONTENT ANALYSIS OF RECENT SURVEY INSTRUMENTS USED IN TURNOVER RESEARCH

Victor H. Appel

University of Texas at Austin

It is axiomatic that every psychological process is ultimately defined by the instrument used to measure it. By this standard there can be little question that the definition of turnover has been undergoing a dramatic change in just the past five years. Since 1976 we have seen remarkable advances in the efforts to devise measures to predict stay/leave behavior or intent both within military and civilian organizations. It will be the intent of this paper to outline some of the major changes that have been made, to suggest additional refinements and to indicate what effects current instrumentation may have on turnover research in the future.

Major Changes in Instrumentation Since 1976

Perhaps the most striking change in turnover questionnaires has been the formulation of theoretically based instruments. Research investigators have heeded the oft-repeated criticism levelled at earlier measures that they lacked an adequate conceptual basis (Porter and Steers, 1973; Price, 1977; and Mobley et al., 1979). As an indication of the theoretical development taking place, one can point to no less than five conceptual models (with corresponding instrumentation) advanced since 1977. Mobley, Horner and Hollingsworth's (1978) Intermediate Linkage Model; Bluedorn's (1979) Path Model of Turnover Intent; Koch and Steers' (1978) Organizational Commitment Model; Martin's (1979) Causal Model of Intent to Leave and Price and Mueller's (1981) Causal Model. Indeed, the expected synthesizing process in which several of the models are combined, has already begun (Bluedorn, 1980).

It would be instructive to compare the best performance achieved by a questionnaire measure tapping dimensions of a model with the best of measures with no specific conceptual basis. Such a comparison might be made between the questionnaire instrument developed from Bluedorn's Causal Model (1978) and the extensively refined Occupational Attitude Inventory (OAI) as developed by the Air Force over the past eight years (Tuttle, Gould and Hazel, 1975; Finstuen, Weaver and Edwards (1981). Using a sample of Army officers, Bluedorn (1979) was able to account for 65% of turnover-intent variance. By contrast, using first term enlisted Air Force personnel, Finstuen, Weaver and Edwards (1981) were able to account for between 46% and 52% of

reenlistment intent variance with a 1973 sample of over 1,000 airmen. Upon replication in 1975, that figure shrank appreciably to 33% to 35% respectively , using a similar sample of over 4,000 airmen. Plainly, the potential of conceptually based questionnaires is promising.

Nature of the Conceptual Models Underlying Instrument Development.

What can be said regarding the types of conceptual models used as the basis for questionnaire construction? Looking at recent measures, are there areas of commonality? There are indeed. Three major attributes will be discussed: 1) Use of the desirability of the alternative construct 2) Adoption of a process focus 3) Incorporation of a wider array of variables, reflecting more comprehensive models.

As indicated by Watson and Appel (1982) a growing consensus has emerged among research investigators in the last several years that the prediction of turnover should be viewed from a systems perspective in which focus is shifted from assessment of level of satisfaction within a single setting to the assessment of the relative satisfaction one has with an existing setting as compared with that of alternative options. This approach recognizes that reenlistment intent or stay/leave behavior should be construed as a decision-making situation in which the individual weighs prospects within an existing context against comparable opportunities perceived to be available elsewhere.

The utility of such a concept as desirability of the alternatives has already been strongly supported by the empirical data. In fact, it has sometimes been found to be the most effective single predictor of turnover. Price and Mueller (1981) stress that that opportunity (their term for the construct) was four times as potent a predictor as pay, the other widely supported predictor. Bluedorn (1978) also found environmental pull (his equivalent term) to be his most potent predictor. These results are even more striking when one considers the wide variability in the ways this construct has been operationalized within survey measures. For example, Schneider (1976) used a full thirty items to compare and contrast the perceived relative efficacy of the Navy with a civilian alternative. Respondents were asked to indicate what they thought their chances were (extremely poor to extremely good) of finding desirable job attributes in the navy or in a civilian context. Variables being compared ranged from "having supervisors who take an interest in you" to "learning new skills and abilities on your job" to "developing close friendships with the people you work with." This extensive, point-by-point comparison contrasts sharply with with more recent measures which use fewer items, and seek less specific bases for assessment.

This difference is most keenly illustrated by examining Price and Mueller's (1981) four item scale to tap the desirability of the alternative variable. They ask the respondent to indicate how easy or difficult they believe it would be to find suitable employment elsewhere. Respondents evaluate how easy or difficult they feel it would be to find employment in their field of endeavor, and of comparable quality to their present position. They are also asked to assess how they perceive the demand within their specialty. Note that these are global items which ask for overall judgments. Still, a third means of operationalizing the construct is illustrated by the DOD Personnel Survey (Undated). Items focus rather specifically on particular aspects of the job, but there are many fewer attributes appraised (only 12). A common operational definition for desirability of the alternatives would allow direct comparison of outcomes achieved.

The second commonality among recent conceptual models deals with the use of a process focus. There is considerable support for the merit of this approach. Both Greenhalgh (1980) and Mobley (1982) argue that previous conceptualizations of turnover have viewed it as a static rather than a dynamic process. A given employee, they argue, is not influenced to the same degree by the same factors at different points during his career. Thus, pay may be relatively more important early in one's career and security increasingly central as one achieves greater tenure. As a consequence, administration of a survey instrument at a single point in time is less descriptive of the actual variability in importance of the variables in question than if a longitudinal study were carried out. In that case, it would permit the use of repeated measures which could reveal the relative importance of a factor over time; that is, at $Time_1$, $Time_2$, $Time_3$ and $Time_4$.

As an example of the process model focus, the Navy is now surveying reactions of first term enlisted personnel to their military experience. This is done at four points in time: 1) While still in boot camp (New Recruit Questionnaire); 2) After completing boot camp (End of Recruit Training Questionnaire); 3) While taking apprenticeship (Class A) training (Training Experience Questionnaire and 4) While at the first duty assignment (Fleet Experiences Questionnaire). Nested questions permit assessment of the salience of differing factors on turnover behavior at each of the four points in time. The use of Process Models can be expected to increase in the future.

The third commonality among recent conceptual models is the increasing scope and complexity of the models. Content has expanded from a near complete focus on job satisfaction/commitment related variables to a much wider coverage. The more recent models embrace variables which are at the economic and organizational level of analysis in additional to the variables at the individual

level. This is readily apparent from even a cursory perusal of the expanded model of the employee turnover process as devised by Mobley, Griffeth, Hand and Meglino (1979). Similarly, one can expect to encounter an increasingly extensive set of demographic, background variables. As reflected by measures used in testing Bluedorn's (1980) unified model, one may expect an increasing array of items used to assess a series of dependent measures. Not only may the traditional turnover intent variable appear in increasingly sophisticated forms, but so too may one find extensive measures of the commitment variable such as the 15 item, Military Career Commitment instrument developed by Butler and Bridges (1978). One may also see items to assess other dependent measures such as the extent of Job Search. Remarkably, these multiple variables are being assessed parsimoniously through the use of Path Analysis techniques (Bluedorn, 1980).

Needed Improvements in Future Turnover Instruments

Despite the significant development taking place in recent turnover measures as outlined above, there are a number of refinements which would be fruitful. Some reference has already been made above to the utility of increased standardization in the operational definitions of key variables. At present there is little consistency across investigators in the items used. Clearly, replication is made more difficult as turnover studies use differing bases to measure Intent to Leave, Desirability of the Alternatives, Job Satisfaction and the like. As a case in point, two relatively sophisticated organizational commitment measures have been derived recently. Mowday, Steers and Porter (1979) have devised their own, 13 item, Occupational Commitment Questionnaire (OCQ), and Gould and Penley have developed a 15 item instrument which they call the Involvement Questionnaire (IQ) (1982). The IQ is based on Etzioni's three types of commitment.

It is becoming increasingly clear that different variables may be predictive of one but not necessarily other turnover outcomes. The variables which tap Intent to Stay/Leave most reliably, may not be the same variables which tap job satisfaction. As a result, research investigators need to clarify for themselves the particular outcome with which they are most concerned, and then construct measures that contain those items pertinent for that purpose. In their recent study, Finstuen, Weaver and Edwards appear to view the Occupational Attitude Inventory as equally applicable to assess job satisfaction, Intent and actual Turnover behavior.

There needs to be increasing specification of the type of person leaving an organization. From which subset are those who leave organizations coming. The recent use of Performance indices may be an important step in addressing this issue. We need clearer specification whether we are losing our best or our marginal personnel.

The critical importance of turnover data for personnel planning argues that turnover research needs to become a continuing part of on-going, organizational information processing systems. This is in contrast to the more typical, current view that periodic turnover studies ought to be conducted. The military services have made much more progress in this regard than has the civilian sector.

Traditionally, studies of turnover have sought to isolate single, generic factors associated with the turnover of all personnel within an organizational setting. Obviously, it is appealing to try to identify a determinant, such as Pay, which would have universal or near-universal relevance to persons at varying levels and across a wide array of occupational groupings. It is already apparent that predictive efficiency may be enhanced as those occupational groups are examined separately. For example, Watson (1982) reports AFHRL efforts at developing AFS-specific regression equations to predict turnover potential.

Implications for the Future

In summary, it is apparent that substantial progress has been made in the last five years to conceptualize the process of turnover. While it is clear from Mobley's (1982) most recent critical analysis that much further work remains, the available measures of turnover stand a much better chance of capturing the complexity of the turnover process than has been possible heretofore.

References

Bluedorn, A. C. A taxonomy of turnover. Academy of Management Review, 1978, 3, 647-651.

Bluedorn, A. C. Structure, environmnent, and satisfaction: Toward a causal model of turnover from organizations. Journal of Political and Military Sociology, 1979, 7, 181-207.

Bluedorn, A.C. A unified model of turnover from organizations. Proceedings: Academy of Management, 40th Annual Meeting of the Academy of Management, Detroit, Michigan, August 9-13, 1980.

Butler, R. P. & Bridges, C. F. Prediction of officer retention prior to commissioning. Journal of Occupational Psychology, 1978, 51, 177-182.

Finstuen, K., Weaver, C. N. & Edwards, J. O. Jr. Predictions of job satisfaction, reenlistment intent, and reenlistment behavior with the Occupational Attitude Inventory. AFHRL -TR-81-XX. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, December, 1981.

Greenhalgh, L.A. A process model of organizational turnover: The relationship with job security as a case in point. Academy of Managcment Review, 1980, 5, 229-303.

Koch, J. L. & Steers, R. M. Job attachment, satisfaction, and turnover among public sector employees. Journal of Vocational Behavior, 1978, 12, 119-128.

Martin, T. N. Jr. A contextual model of employee turnover intentions. Academy of Management Journal, 1979, 22, No. 2, 313-324.

Mobley, W. H. Intermediate linkages in the relationship between job satisfaction and employee turnover. Journal of Applied Psychology, 197-, 62, 237-240.

Mobley, W. H., Horner, S. O. & Hollingsworth, A. T. An evaluation of precursors of hospital employee turnover. Journal of Applied Psychology, 1978, 63, No. 4, 408-414.

Mobley, W. H., Griffeth, R. W., Hand, H. H. & Meglino, B. M. Review and conceptual analysis of employee turnover process. Psychological Bulletin, 1979, 86, 493-522.

Mobley, W. H. Some unanswered questions in turnover and withdrawal research. Academy of Management Review, 1982, 7, 111-116.

Mowday, C., Steers, R.M., & Porter, L.W. The meaurement of organizational commitment. Journal of Vocational Behavior, 1979, 14, 224-2

Porter, L. W. & Steers, R. M. Organizational, work and personal factors in employee turnover and absenteeism. Psychological Bulletin, 1973, 80, 151-176.

Price, J. L. The Study of Turnover. Ames: Iowa State University Press, 1977.

Price, J. T. & Mueller, C. W. A causal model of turnover for nurses. Academy of Management Journal, 1981, 24, 543-565.

Tuttle, T. C., Gould, R. B., & Hazel, J. T. Dimensions of job satisfaction: Initial development of the Air Force Occupational Attitude Inventory. AFHRL-TR-75-1, AD-A014 796. Lackland AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, June, 1975.

Watson, T. W., & Appel, V. H. The desirability of the alternatives: A promising construct for understanding turnover decisions. In Proceedings of the 8th Psychology in the DOD Symposium, Colorado Springs: USAFA, April 1982.

Watson, T. W. Process Models of Turnover Within an Open-Systems Context. In Proceedings of the 1982 Conference of the Military Testing Association, San Antonio, Texas: November 1982.

Vocational Interests and Job Satisfaction: Effects on Turnover Among Air Force Enlistees

Michael D. Matthews
Air Force Human Resources Laboratory
Brooks Air Force, Texas 78235

The Vocational Interest-Career Examination (VOICE) is an Air Force instrument designed to assess vocational interests among Air Force enlistees. Its development and validation are described by Alley and Matthews (1982). In addition, job satisfaction can be predicted by the VOICE (Alley, Wilbourn, & Berberich, 1976). Job satisfaction has been found to be related to fatigue, dissatisfaction with life, depression, psychosomatic illness, mental illness, drug and alcohol abuse, job performance, and coronary heart disease (Cf. Alley & Matthews, in press). Perhaps the most serious implication of personnel dissatisfaction, however, has to do with its influence on various forms of occupational withdrawal. Research has demonstrated quite consistently that personnel dissatisfied with their jobs are much more likely to be absent from their work (Waters & Roach, 1973) and to terminate their employment at a higher frequency than are satisfied workers (Mobley, Griffeth, Hand, & Meglino, 1979).

The diverse and serious implications of job dissatisfaction led the Air Force Human Resources Laboratory to initiate a study of the relationship between vocational interests among first-term enlisted accessions, as assessed by the VOICE, and attrition from the Air Force. Preliminary results from this research program ~~have been presented earlier~~ by Matthews (1982) and Matthews and Berry (1982). The purpose of this paper is to present additional findings from this research program.

Method

Subjects

36,759 male and 12,909 female 1973-1975 Air Force enlisted accessions were administered the VOICE during basic training and tracked through their initial tour of duty. The subjects were typical of past Air Force accessions. Their average age was 18, the racial composition of the sample was similar to that of the United States population as a whole, and most (95.29%) had completed high school.

The VOICE

The VOICE consists of a 300-item vocational interest inventory requiring approximately 30 minutes to administer. Individual items are presented in booklet form and consist of occupational titles, work tasks, leisure time activities, and desired learning experiences. Respondents indicate relative preferences for each item in a standard like-indifferent-dislike (LID) format. Item responses were converted to two types of scales: (a) basic interest scales, and (b) occupational scales. The basic scales represent measures of general interest in various occupational and technical areas. They were constructed by grouping items of similar content into 18 independent sets covering a wide range of interests in the vocational and technical domain. The basic interest scales cover areas of Office Administration, Electronics, Heavy Construction, Science, Outdoors, Medical Service, Aesthetics, Mechanics, Food Service, Law Enforcement, Audiographics,

Mathematics, Agriculture, Teacher/Counseling, Marksman, Craftsman, Drafting, and Automated Data Processing. All items within each scale are homogeneous in the sense that each was selected to measure the same underlying dimension. The Office Administration items, for example, measure interest in clerical, administrative, and business related activities.

The occupational scales were designed for use in evaluating job assignment alternatives. It has been found that certain patterns of basic interest scores predict job satisfaction in various Air Force job clusters (Alley et al., 1976). These clusters, 20 in number, represent an exhaustive categorization of Air Force job specialties. The VOICE occupational scales, therefore, provide a predicted job satisfaction score for each of these 20 job clusters. Consequently, if used operationally job placement personnel would be able to readily obtain a prediction of job satisfaction for any Air Force career field, by determining in which of the clusters that particular job falls . The occupational scales, while formulated from basic interests, provide direct estimates of job satisfaction for each career field in the set and can be used for making specific comparisons between alternative assignments (Alley et al., 1976). Predicted job satisfaction (PJS) scores range from 200 to 800, with a mean of 500 and standard deviation of 100. For a more thorough and technical discussion of the development of the VOICE and a description of the basic interest and occupational scales, their psychometric characteristics, and validity, see Alley and Matthews (1982).

Procedure

The sample of recruits was monitored until completion of their initial four to six year duty obligation and cumulative attrition rates were assessed after 12, 24, and 36 months of service. Each subject's career field of assignment was identified and the PJS score associated with that field determined. Attrition rates were determined for each of the 20 VOICE DOD occupational clusters, and by sex within clusters. The occupational clusters were then combined for an overall analysis of attrition as a function of PJS score. Finally, these overall data were broken out by sex to examine possible effects of gender on the relationship between PJS scores and attrition. In addition, pre-enlistment variables including age, education level, Armed Services Vocational Aptitude Battery (ASVAB) scores, and Armed Forces Qualification Test (AFQT) scores were obtained for each subject. These variables are known to be related to Air Force attrition rates (Finstuen & Alley, in press) and, together with VOICE PJS scores, were entered into regression models designed to identify the sources and magnitude of variance predictive of attrition.

Results and Discussion

The relationship between predicted job satisfaction and attrition from the Air Force at 12, 24, and 36 months of service is depicted in Figure 1, which presents the percentage of cases lost from the Air Force as a function of predicted job satisfaction. For example, approximately 40 percent of subjects who had low predicted job satisfaction scores had attrited within 36 months of their initial enlistment, versus 26 percent of the group with high predicted job satisfaction scores.

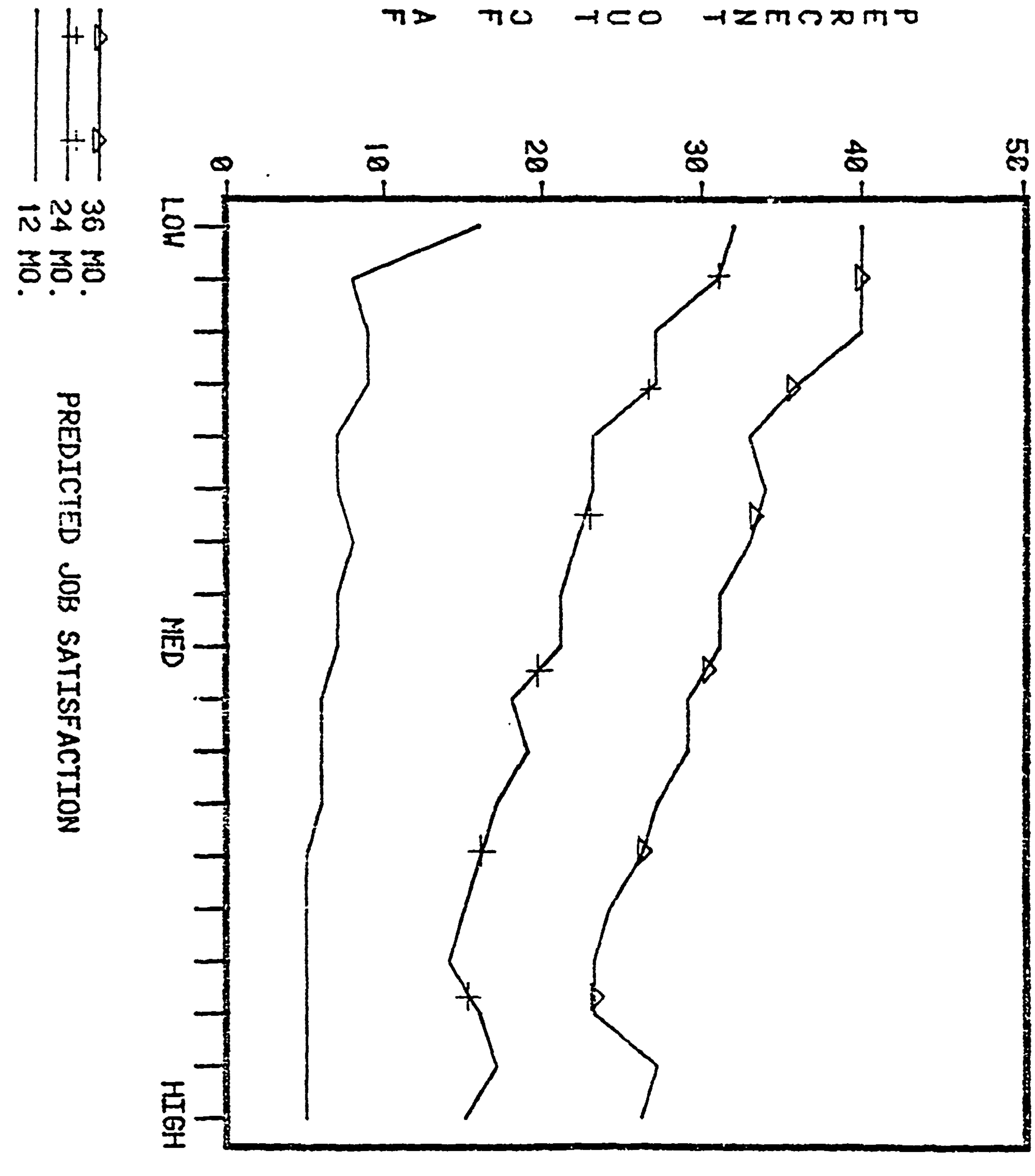


Figure 1. Cumulative attrition rates as a function of predicted job satisfaction, after 12, 24, and 36 months of service.

A regression analysis was conducted on the 36 month attrition rates. A full regression model (n=51,916) containing vectors for age, education level, ASVAB composite scores, AFQT scores, squares of all the above, cubes of all aptitude variables, and VOICE PJS scores was developed. This full model resulted in a significant ($F=50.93$; $df=21, 51,894$; $P < .001$) R of .142. A restricted model, differing from the full model only in the deletion of the PJS vector, also significantly predicted attrition ($F=41.90$; $df=20, 51,895$; $P < 001$) with an R of .126. Moreover, the difference between the Rs of the full and restricted models was also significant ($F=227.80$; $df=1, 51,894$; $P< .001$), indicating that the VOICE predicts attrition above and beyond the influence of other pre-enlistment variables. Finally, VOICE PJS scores alone were significantly related to attrition ($F=334.07$; $df=1, 51,914$; $P < .001$), with an R of .080. An examination of the correlation matrix (not shown) of the pre-enlistment variables and attrition showed that only high school graduation ($r=.088$) correlated higher with attrition than did VOICE PJS scores.

The statistical analysis of the relationship between PJS scores and attrition presented in Figure 1 indicates (1) that PJS scores are significantly related to attrition, and (2) PJS scores add to the predictive power of other pre-enlistment variables, such as age, education level, and aptitude level. It is possible that the small, but significant, relationship between PJS scores and attrition would be more substantial if the analysis differentiated sources of attrition unlikely to be related to predicted job satisfaction (eg., death, disability) from those sources likely to be affected by job satisfaction (eg., marginal performance). Finally, additional regressions testing the effects of the predictor variables on attrition within each of the 20 DOD occupational clusters may reveal a greater or lesser degree of relationship between predictor and criterion variables than did the overall analysis.

In conclusion, the results from the present study indicate a small but reliable relationship between predicted job satisfaction and Air Force enlisted attrition. Data from this study suggest that utilization of VOICE PJS scores in the classification of recruits to career fields would have a major impact on attrition rates with consequent decreases in training costs and an improvement of overall force quality.

References

Alley, W. E. & Matthews, M. D. The Vocational Interest-Career Examination: A description of the instrument and possible applications. Journal of Psychology, 1982, 112, 169-193.

Alley, W. E., Wilbourn, J. M., & Berberich, G. L. Relationships between performance on the Vocational Interest-Career Examination and reported job satisfaction. AFHRL-TR-76-89, AD-A040 754. Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1976.

Finstuen, K., & Alley, W. E. Occupational correlates of first-term enlisted tenure. AFHRL-TR-82-36. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, in press.

Matthews, M. D. Vocational interests, job satisfaction and turnover among Air Force enlistees. Proceedings, Fourth International Learning Technology Congress and Exposition, February 1982.

Matthews, M. D., & Berry, G. A., A preliminary look at vocational interest and attrition among Air Force enlistees. Proceedings, Eight Psychology in the DOD Symposium, April 1982.

Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. Review and conceptual analysis of the employee turnover process. Psychological Bulletin, 1979, 86, 493-522.

Waters, L. K., & Roach, D. Job attitudes and predictors of termination and absenteeism: Consistency over time and organizational units. Journal of Applied Psychology, 1973, 87(3), 341-342.

Definition, Classification, and Measurement of Turnover in Civilian and Military Contexts

Thomas W. Watson

Air Force Human Resources Laboratory

Employers invest considerable resources in recruiting and training employees, and the loss of personnel can be costly. Losses also make it difficult for organizations to develop a career force with levels of experience and proficiency necessary for optimum organizational effectiveness. Thus, employers have focused attention on turnover.

Much turnover research has been conducted, primarily to determine factors influencing termination or retention decisions. Although knowledge has been accummulating for decades, certain issues have been overlooked. For instance, little attention has been focused on the criterion issue. Considerably more effort has been directed toward identifying the antecedents of turnover, than in carefully defining or classifying criteria, or in identifying optimal measurement methods for specific purposes. The effort invested in the predictor set has been productive, but more consideration needs to be directed toward termination/retention criteria and methods of measurement.

This paper examines the termination/retention criterion issue. First, the need for precise definition, classification and measurement will be discussed; then, attempts to define and classify turnover and retention in military and civilian contexts will be presented; finally, measurement will be discussed.

The Need for More Precise Definition, Classification, and Measurement

The distinction between retention and turnover appears easy to specify: turnover occurs if an employee leaves, and retention occurs if employment is continued. However, turnover is sometimes broadly defined to include entry into as well as exit from an organization. Also, many conditions exist under which turnover can occur. For instance, an incumbent may leave voluntarily, or may be forced to leave involuntarily. These categories can be further subdivided. Likewise, employment might be terminated prior to fulfillment of a contractual obligation. Separation under such circumstances is defined differently than separation under conditions of no obligated term of service. Even retention can be subcategorized. For instance, those who remain can be differentially classified on the basis of productivity or commitment. Thus, definition, classification, and measurement of these complementary terms is more difficult than would initially appear to be the case.

Understanding of turnover and retention can be enhanced by examining how turnover has been defined and classified. If more precise subcategories can be developed in which different types of stayers or leavers are not inadvertently grouped together for research or applied purposes, error in prediction equations can likely be reduced. The variables which influence specific types if turnover can also better be identified. Such information will also provide more precise information for making management decisions.

Another important consideration exists. A stay/leave criterion often is not available, or may not be the most desirable criterion to use. An interim

or surrogate dependent measure such as behavioral intent is frequently used. For some purposes, such a criterion is preferred, for it gives management an opportunity to identify problems and take remedial action to induce valued employees to stay. The definition, classification and measurement of surrogate criteria also need to be examined.

Definition and Classification of Separation and Retention Criteria

To measure a construct and use it effectively it needs to be carefully defined, and tailored to the intended use. Without precise definition, the validity of measures can be questioned. In addition, if a multifaceted construct like turnover is defined and measured as if it were unidimensional, its utility will be diminished. It is the purpose of this section to examine how turnover has been defined and classified. While broad definitions of turnover are of limited practical utility, they can provide a starting point for understanding turnover and developing a refined classification. Selected broad definitions are discussed below.

Definition: Defining Turnover in Its Broadest Sense

Broad Definitions From the Civilian Literature. Belknap (1977) provided a straightforward macro-definition of turnover that conveys the implicit definition most researchers and practioners hold. "Turnover is anyone who [leaves the company]" (p. 233). Likewise, Forrest, Cummings, and Johnson (1977) noted that some definitions of turnover include all leavers while others exclude personnel who leave for particular reasons, and advocated the more global approach. Based on a skepticism concerning reasons for leaving (see Lefkowitz & Katz, 1969), they recommended the following definition: removal from an organization's payroll. Mobley (1982) provided a similar definition: cessation of organizational membership by someone receiving compensation from the organization.

The definitions provided by Belknap (1977) Forrest el at. (1977) and Mobley (1982) focus on all who leave an organization. However, leaving a job might include movement within a firm and, as Van der Merwe and Miller (1975) pointed out, most authors do not consider such movement to be turnover. Carr (1972) took exception to the practice of excluding intraorganization transfers and argued that turnover includes the continuing movement of employees within organizations. He defined such movement as internal turnover, and used the term total turnover to describe the flow of human resources into, through, and out of an organization. In this second respect, Carr's definition is similar to the macro-definition proposed by Price: "the degree of individual movement across the membership boundaries of a social system" (Price, 1977, p. 4). Price (1975) stated that movement included the entry and exit of individuals. Unlike Carr (1972), he excluded intraorganizational transfers or promotions.

Although Price (1977) deviated from common practice by including both inward and outward migration in his definition of turnover, he favored a defintion of turnover also favored by most other social scientists: movement of individuals. This is in contrast to turnover as an aggregate phenomenon often used by economists. Bluedorn (1978) also described turnover as representing a change in an individual's membership status, but noted problems with Price's conceptualization, since directionality was not specified. He stressed that turnover commonly refers to the act of leaving an organization, but described turnover as leaving in a narrower sense than did Belknap (1977),

Forrest et al. (1977), or Mobley (1982) by emphasizing that turnover usually refers to people who quit, not to people who are fired.

Military Approaches to Defintion. Most of the broad definitions of turnover discussed thus far have been developed in the civilian literature. Turnover is seldom used in reference to the loss of military personnel, but the military tends to define termination in a similar fashion to the civilian sector. Organizational exit and termination of membership are emphasized. For example, AFR 35-41 and DOD Directive 1332.14 both define discharge as the termination of enlistments (or appointments) or other military status. However, despite similarities, there are also differences. Military separations are defined in a broader sense than turnover in the civilian community, referring to a change in status which may involve termination, or retention. The military also provides a far more elaborate subclassification of reasons for separation. Another important difference exists. Due to the unique character of the military employment contract, which usually involves a period of obligated service, separation prior to completion of a contractual active-duty obligation is referred to as attrition (see DOD Directive 1415.7). Although this type of turnover may apply to selected civilian jobs, it is infrequently used in such contexts. Turnover defined as attrition is usually categorized in terms of when it occurs during a military member's term of obligated service.

Classification: Developing a Taxonomy of Turnover

Attempts at classifying turnover into a variety of subcategories represents a shift in focus from broad definition to the development of a taxonomy of turnover. Using Bluedorn's (1978) efforts to develop a taxonomy as a starting point, this topic is discussed below.

Voluntary vs Involuntary Turnover. Bluedorn (1978) developed a taxonomy of turnover based on two dimensions: (1) the direction of movement across the organizational boundary (i.e., in or out), and (2) the source of initiation of movement (i.e., the individual (voluntary), or by forces other than the individual (involuntary). Like Bluedorn, Price (1975, 1977) subcategorized turnover into voluntary and involuntary categories, placing primary emphasis on voluntary turnover. Price defined voluntary turnover as "individual movement across the membership boundaries of a social system which is initiated by the individual" (Price, 1977, p. 9), and defined involuntary turnover as movement not initiated by the individual.

Controllable vs Uncontrollable Turnover. A distinction has also been made between controllable and uncontrollable turnover. Controllable and uncontrollable turnover, as defined by Van der Merwe and Miller (1971), appear to be synonomous with avoidable and unavoidable turnover, which are terms also found in the literature. Van der Merwe and Miller (1971) defined controllable turnover as the avoidable loss of personnel since management action could have been taken to reduce, or prevent, the loss. The rationale for emphasis upon controllable turnover, according to Van der Merwe and Miller (1971, 1975) is based on the premise that any approach to the measurement of turnover used for management decision making should distinguish between turnover which is within management control from that which is not. There is, however, disagreement concerning what is under management control. Van der Merwe and Miller proposed that employee-initiated separations (i.e., voluntary separations) and employer-initiated separations (i.e., dismissals) be included under the heading of controllable turnover. They recommended against further

differentiation on the basis of reasons for leaving due to the unreliability of statements made by employees at the time of exit. Van der Merwe and Miller (1971, 1975) included voluntary separations and dismissals in their definition of controllable turnover. Other's have defined controllable turnover in a narrower sense. Price (1977) noted that controllable turnover is often similar in meaning to voluntary turnover.

Other investigators have also recommended subcategorization of turnover into categories similar to those proposed by Van der Merwe and Miller (1971, 1975). Dalton, Krackhardt, and Porter,(1981), and Dalton, Tudor, and Krackhardt (1982) recommended that voluntary turnover be subdivided into unavoidable and controllable categories. Likewise, Lefkowitz and Katz (1969) recommended that turnover be subcategorized as involuntary, avoidable voluntary, and unavoidable voluntary.

Functional vs Dysfunctional Turnover. Historically, emphasis has been placed on voluntary turnover. In addition, turnover has been construed to have negative consequences for both individuals and organizations. More recently, Dalton and his colleagues (Dalton, Krackhardt, & Porter, 1981; Dalton & Tudor, 1982; Dalton, Tudor, & Krackhardt, 1982) have challenged the notion that voluntary turnover is invariably detrimental to organizations. They have taken the classification of turnover on the basis of positive versus negative organizational consequences and developed this theme into an expanded taxonomy of turnover emphasizing the distinction between functional and dysfunctional turnover. They focused on the incumbent's evaluation of the organization and the organization's evaluation of the incumbent. They also expanded upon the traditional approach by looking at high-quality and low-quality employees and the organizational outcomes associated with voluntary turnover among these two different groups of employees.

According to Dalton, Krackhardt, and Porter (1981) dysfunctional turnover occurs when an individual wants to leave the organization, and the organization desires to retain the individual. In contrast, functional turnover occurs when an individual wants to leave but the organization is unconcerned, due to a negative evaluation of the individual. Such turnover is considered beneficial to the organization. Dalton and his associates included the criteria of employee quality and replaceability in their exposition of functional/dysfunctional turnover, and applied these criteria to stayers and leavers. Implicit in a quality dimension is performance. Other investigators have also emphasized the importance of classifying those who leave along a performance dimension. For instance, Porter and Steers (1973) recommended that more attention be given to the turnover of differentially valued employees. Martin, Price, and Mueller (1981) noted that most of the relevant literature indicates that incumbents who perform better are most likely to leave. This finding is consistent with the importance of alternative job prospects stressed by Price (1977) and by Watson and Appel (1982). Higher performers would be most likely to leave since they would have the greatest employment opportunities external to their present work environment.

Military Efforts at Precise Classification of Reasons for Separation. Generally, civilian efforts toward classification have yielded the rather broad categories previously discussed. The DOD departments have developed an elaborate system for coding reasons for leaving. Separation Program Designator codes are assigned to those separating from the service, or transferring between services. These three-digit codes represent separation type and separation reason. However, until recently, considerable variation

existed in the separation classification schemes developed by the different services. In recent months, efforts have been underway to increase the uniformity of classification across the DOD. This has been in response to DOD Directive 1332.14, dated 28 January, 1982, which applies to administrative separations after 1 October 1982. This directive provides current guidance concerning the manner in which the military services classify separations. The major subcategories of separations contained in this directive are as follows: (a) Expiration of Service Obligation, (b) Selected Changes in Service Obligation, (c) Convenience of the Government, (d) Disability, (e) Defective Enlistments and Inductions, (f) Entry Level Performance and Conduct, (g) Unsatisfactory Performance, (h) Homosexuality, (i) Drug Abuse Rehabilitation Failure, (j) Alcohol Abuse Rehabilitation Failure, (k) Misconduct, (l) Separation in Lieu of Trial by Court-Martial, (m) Security, (n) Unsatisfactory Participation in the Ready Reserves, (o) Secretarial Plenary Authority, and (p) Reasons Established by the Military Department. The Defense Manpower Data Center provides another useful taxonomy referred to as Interservice Separation Codes, as follows: (a) Release form Active Service, (b) Medical Disqualifications, (c) Dependency or Hardship, (d) Death, (e) Entry into Officer Programs, (f) Retirement (Other than Medical), (g) Failure to Meet Minimum Behavioral or Performance Criteria, (h) Other Separations or Discharges.

The Measurement of Separation and Retention

As Baysinger and Mobley (1982) noted, turnover and retention are individual and aggregate phenomena. As an individual phenomenon, turnover is frequently measured using an intent criterion, or as a dichotomous stay/leave criterion. As an aggregate phenomenon, measures such as accession and separation rates, stability and instability rates, and survival and wastage rates are computed. Space limitations preclude discussion of how these rates are computed For detailed accounts, the reader is referred to Van der Mere and Miller (1975) and Price (1977). Selection of individual or aggregate measures often varies with intended use or professional specialization. Although there are exceptions, economist and practitioners interested in human reources management frequently use aggregate measures while psychologists are more inclined to use individual measures.

In light of issues raised earlier in this paper, it is important that researchers and managers consider how they wish to define and classify turnover before developing a measure appropriate for their intended use. The question of what type of organizational movement constitutes turnover should be addressed first. Next, consideration should be given to how one wishes to subclassify the broader set of employees under study. After defining and classifying the target population of interest, the investigator should then decide on a method of individual or aggregate measurement which is tailored to the target population of interest, and the intended use.

Summary and Conclusions

Turnover and retention are complex phenonema. Researchers need to consider the interelationship of organizational entry, intraorganizational movement, and organizational exit. They also need to consider how best to define, classify, and measure turnover and retention on the basis of particular research or applied needs. In addition to identifying the relative contribution of determinants of turnover, more consideration needs to be given to the dependent variable itself.

References

Baysinger, B. D., & Mobley, W. H. *Employee turnover: Individual and organizational analyses*. TR-ONR-1, ADA 114 439. Arlington, VA: Office of Naval Research, April 1982.

Belknap, W. C. Turnover - One company's approach. In H. W. Sinaiko (Ed.), *First term enlisted attrition - Vol 1: Papers*. TR-3 Arlington, VA: Office of Naval Research, June 1977.

Bluedorn, A. C. A taxonomy of turnover. *Academy of Management Review*, 1978, *3*, 647-651.

Carr, C. B. Total turnover. *Personnel Journal*. 1972, *51*, 524-527.

Dalton, D. R., Krackhardt, D. M., & Porter, L. W. Functional turnover: An empirical assessment. *Journal of Applied Psychology*, 1981, *66*, 716-721.

Dalton, D. R., Todor, W. D., & Krackhardt, D. M. Turnover overstated: The functional taxonomy. *Academy of Management Review*, 1982, *7*, 117-123.

Forrest, C. R. Cummings, L. L., & Johnson, A. C. Organizational participation: A critique and model. *Academy of Management Review*, 1977, *2*, 589-601.

Lefkowitz, J., & Katz, M. L. The validity of exit interviews. *Personnel Psychology*, 1969, *22*, 445-455.

Martin, T. N., Price, J. L., & Mueller, C. W. Job performance and turnover. *Journal of Applied Psychology*, 1981, *66*, 116-119.

Mobley, W. H. *Employee turnover: Causes, consequences, and control*. Reading, MA: Addison - Wesley, 1982.

Price, J. L. A theory of turnover. In B. O. Pettman (Ed.), *Labour turnover and retention*. New York: John Wiley & Sons, 1975.

Price J. L. *The Study of turnover*. Ames: Iowa State University Press, 1977.

Van der Merwe, R., & Miller, S. The measurement of labour turnover. *Human Relations*, 1971, *3*, 233-253.

Van der Merwe, R., & Miller, S. The measurement of labour turnover. In B. O. Pettman (Ed.), *Labour turnover and retention*. New York: John Wiley & Sons, 1975.

Watson, T. W., & Appel, V. H. The desirability of alternatives: A promising construct for understanding turnover decisions. In *Proceedings of the 8th Psychology in the DOD Symposium*, Colorado Springs: USAFA, April 1982.

Process Models of Turnover Within an Open-Systems Context

Thomas W. Watson

Air Force Human Resources Laboratory

Employers are interested in retaining experienced, productive personnel. Thus, employee turnover has been a topic of concern. During the past decade, there has been renewed interest in turnover, as documented in a recent bibliography (Berry, Weaver, Watson, & Finstuen, Note 1) and in recent literature reviews (see Porter & Steers 1973; Hand, Griffeth, & Mobley, 1977; Price, 1977; Mobley, Griffeth, Hand, & Meglino, 1979; and Muchinsky & Tuttle, 1979).

start

Despite renewed conceptual and empirical interest, researchers have been only moderately successful in enhancing our ability to understand or predict turnover behavior. Both Greenhalgh (1980) and Mobley (1982) have lamented that our understanding of turnover remains limited. They proposed that our understanding could be advanced by examining turnover from a process perspective and recommended that future turnover research incorporate such a perspective. Watson and Appel (1982) noted that until recently even the most sophisticated studies were accounting for no more than 25% of the turnover variance. They provided theoretical and empirical support for a desirability-of-alternatives construct and advocated a shift in conceptual focus. Watson and Appel recommended that future research on turnover gather comparative data concerning incumbent perceptions of prospects in existing work settings vis a vis desirable and obtainable alternative work contexts. In so doing, they advocated an open-systems approach to the study of turnover which is compatible with the process perspective advocated by Greenhalgh (1980) and Mobley (1982).

The Air Force is concerned about factors influencing separation and retention decisions, and the Air Force Human Resources Laboratory (AFHRL) has been involved in retention research for several years. One component of this program involves process models of turnover research, which is designed to correct the shortcomings noted by Greenhalgh (1980), Mobley (1982), and Watson and Appel (1982) by examining turnover from both process and open-systems perspectives. It provides new directions for turnover research which should enhance our understanding of turnover decisions, and increase the proportion of explained variance in turnover. This paper provides a brief description of AFHRL process models research, followed by a discussion of the process and open-systems perspectives, as well as the manner in which these perspectives have been incorporated in the process models research design.

The process models turnover research at AFHRL involves a longitudinal assessment of factors influencing separation and retention decisions of Air Force enlisted personnel. An initial survey instrument is currently under development for longitudinal administration to a random sample of approximately 15,000 first- and second-term Air Force enlisted personnel. At the time of initial survey, these participants will be at different points in their career, and will be surveyed yearly for at least three consecutive years to identify changes in the factors influencing their intent to stay or to leave over time. To provide a check on the validity and reliability of the information obtained, and to gain a richer understanding of the factors

influencing separation and retention decisions, a subset of participants will also be surveyed via phone. Occupation-specific and career-cycle-specific models of the turnover decision process will be developed and tested, initially using a behavioral-intent criterion. Actual stay/leave behaviors will be determined later via personnel data files, and used as an additional dependent measure. Multivariate analysis techniques such as regression analysis, path analysis, and discriminant analysis will be used to assess the relative influence of the variables examined on turnover decisions.

The Process Perspective in the Process Models Research

Historically, turnover has been regarded as a static phenomenon. A static view is conceptually simple and convenient since it does not require consideration of change over time and requires only a single time assessment. However, this view appears to be an oversimplification which ignores the dynamic nature of the turnover process. Mobley (1982) criticized traditional turnover research for neglecting the critical feature of turnover as a process: change over time. In his judgement, repeated measures of multiple factors over time are essential to understand this process better. He further cautioned that multivariate analyses need to have a strong conceptual base in order to enhance our understanding of the process of turnover.

The AFHRL process models research incorporates a process perspective since turnover is viewed as a dynamic process characterized by change over time. By using multiple surveys over a period of years, changes in the relative weight of turnover determinants over time can be assessed. Such an approach can also better illuminate the stepwise character of turnover decisions.

A longitudinal approach is being taken because change over time is so critical to a process perspective. Although all process models of turnover stress this characteristic, some investigators, such as Mobley (1977), Mobley, Horner, and Hollingsworth (1978), and Steers and Mowday (1981) focus primarily upon changes occuring over a relatively short period of time. For instance, Mobley and his associates have proposed an intermediate-linkage model of the turnover decision process. This is a multi-step model involving thoughts of quitting, intention to search, actual search, and the intention to leave as orderly, sequential precursors to the act of leaving. Other investigators, such as Greenhalgh (1980) have described process models which encompass the entire career cycle. Greenhalgh's model is based on the assumption that the decision process concerning staying or leaving should be traced throughout a person's career with an organization. At different points in an individual's career cycle, different factors become salient. Thus, changes in the factors influencing stay/leave decisions need to be examined over extended periods of time.

In the process models research conducted by AFHRL, a survey instrument is currently being developed to measure the dynamic nature of the turnover process over the short term, and as individuals progress in their careers. Information will be gathered concerning the intermediate steps which precede the actual stay/leave decision. Information will also be obtained on multiple factors presumed to influence turnover decisions and to have a differential impact over time. Variables considered for inclusion in the survey instrument are being selected on the basis of their theoretical or empirical relationship to turnover decisions. Air Force managers concerned with retention issues have also been consulted to identify variables uniquely applicable to life in

the Air Force. Variables are being selected with care, and in accord with Mobley's (1982) recommendation, the multivariate approach is conceptually and empirically based. Promising factors considered for inclusion (other than biodemographic variables) are as follows: absolute and relative job satisfaction; satisfaction with the Air Force way of life; pay and fringe benefits; family responsibilities and the attitudes of family members; frequency of, and satisfaction with assignments; met expectations; promotion opportunities; organizational commitment and behavioral intent; immediate affective and behavioral precursors to behavioral intent; the desirability and availability of alternative employment opportunities; transferability of skills to the civilian sector; and the perceived utility of Air Force versus civilian employment. Additional variables, like biodemographic data, and information pertaining to economic conditions, will be obtained from other sources such as AFHRL personnel data files or the Bureau of Labor Statistics. The rationale for variable selection is discussed further in the following section on the open-systems perspective. This is not intended to be an exhaustive list.

The Open-Systems Perspective in the Process Models Research

In addition to surveying respondents longitudinally to determine changes in the factors influencing separation and retention decisions over time, the AFHRL process models research is designed to examine turnover within an open-systems context. An open-system is one that influences, and is influenced by other systems. This perspective acknowledges that organizations are embedded in a larger social context and that factors in the external environment can have an impact on turnover decisions. Most psychologically-oriented turnover research has not taken such a perspective. Rather, most such research has taken an implicit closed-systems perspective wherein intraorganizational determinants have been emphasized, and the impact of external factors has been ignored.

Some investigators have taken, or advocated, a more open-systems approach. For instance, economists have long been interested in factors such as unemployment rates. Family responsibilities and the impact of the desires of one's spouse have been given some attention. Consideration of the impact of alternative work contexts was advocated as early as the 1950's and 1960's by March and Simon (1958), and Smith, Kendall, and Hulin (1969). However, these authors' recommendations were largely ignored, and not until recently has the impact of desirable and obtainable alternatives been given serious consideration. This construct, which Watson and Appel (1982) called the desirability of alternatives, has been given a variety of names and been operationally defined in a variety of ways. However, considerable empirical support for the importance of this construct as a determinant of turnover has been accumulating (see, for example, Schneider, 1976; Bluedorn, 1979; and Price and Mueller, 1981a, 1981b).

Bronfenbrenner (1979) has exposited a theory supporting an open-systems approach. Although Bronfenbrenner (1979) was interested in the ecology of human development, his work can readily be applied to turnover research. He portrayed the individual as interacting with an ecological environment conceived as a set of nested structures, defined in terms of the following interrelated systems: the microsystem, the mesosystem, the exosystem, and the macrosystem. He advocated that the possible impact of all these interrelated systems on the behavior of individuals needs to be considered.

Relating Bronfenbrenner's conceptualization to turnover, the microsystem refers to the work environment in which the individual participates. The mesosystem refers to other social systems in which he or she also participates, such as a family. The exosystem refers to alternative work settings in which the individual does not necessarily participate but about which he or she has knowledge, and which can influence his or her behavior. The macrosystem refers to the society in which the individual lives. From an open-systems perspective, data concerning all of these systems would need to be collected. Thus, in addition to measuring perceptions of an incumbent's current work setting (the closed-systems approach), information such as the attitudes of family members, perceptions of alternative work settings, and economic conditions in the society at large would need to be gathered.

As the partial list of variables provided earlier in this paper suggests, the AFHRL process models research incorporates an open-systems perspective. Not only will information be gathered concerning an incumbent's existing work environment, relative perceptions of future propects in one's existing setting compared with desirable and obtainable alternative work contexts will also be measured. This approach measures relative satisfaction in multiple work contexts and the perceived utility of multiple work environments for the attainment of desired outcomes. Not only will comparative perceptions of the microsystem and alternative exosystems be obtained, so also will information about to the mesosystem. For instance, the impact of family size and responsibilities will be assessed, along with the impact of career aspirations of one's spouse, or extended separations from family members during unaccompanied assignments. The macrosystem will also be examined. Taking an interdisiplinary approach, data will be gathered on the state of the economy and the extent of employment opportunities in the external environment.

Summary

In two major respects, the AFHRL process models investigation provides new directions for turnover research which hold promise for improving our ability to understand and predict separation/retention decisions. First, this research provides for assessment of the dynamic nature of turnover via longitudinal survey administration. Second, the research goes beyond the traditional assessment of intraorganizational factors by also considering the impact of factors in the external environment. Thus, it incorporates both a process perspective and an open-systems perspective.

The research is innovative in other ways. For instance it uses a conceptually based multivariate design and multimethod survey techniques. Although the process models research provides promising new directions for turnover research, there are also problems which will require attention. For instance, some of the most promising new variables, such as the desirability of alternatives have not been adequately defined operationally or measured with sufficient precision. Also, as with all survey research, there are problems concerning the reliability and validity of responses, and of nonresponse, which will need to be resolved. The use of both paper-and-pencil and telephone survey techniques should attenuate such problems.

References Notes

1. Berry, G. A., Weaver, C. N., Watson, T. W., & Finstuen, K. A bibliography of military and non-military personnel turnover literature. AFHRL-TP-81-22. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory, in press.

References

Bluedorn, A. C. Structure, environment, and satisfaction: Toward a causal model of turnover from organizations. Journal of Political and Military Sociology, 1979, 7, 181-207.

Bronfenbrenner, U. The ecology of human development. Cambridge, MA: Harvard University Press, 1979.

Greenhalgh, L. A. A process model of organizational turnover: The relationship with job security as a case in point. Academy of Management Review, 1980, 5, 299-303.

Hand, H. H., Griffeth, R. W., & Mobley, W. H. Military enlistment, reenlistment and withdrawal research. A critical review of the literature. TR-3, AD A048955. Arlington VA: Office of Naval Research, December 1977.

March, J. G., & Simon, H. A. Organizations. New York: John Wiley, 1958.

Mobley, W. H. Intermediate linkages in the relationship between job satisfaction and employee turnover. Journal of Applied Psychology, 1977, 62, 237-240.

Mobley, W. H. Some unanswered questions in turnover and withdrawal research. Academy of Management Review, 1982, 7, 111-116.

Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. Review and conceptual analysis of the employee turnover process. Psychological Bulletin, 1979, 86, 493-522.

Mobley, W. H., Horner, S. O., & Hollingsworth, A. T. An evaluation of precursors to hospital employee turnover. Journal of Applied Psychology, 1978, 63, 408-414.

Muchinsky, P. M., & Tuttle, M. L. Employee turnover: An empirical and methodological assessment. Journal of Vocational Behavior. 1979, 14, 43-77.

Porter, L. W., & Steers, R. M. Organizational, work, and personal factors in employee turnover and absenteeism. Psychological Bulletin, 1973, 80, 151-176.

Price, J. L. The study of turnover. Ames: Iowa State University Press, 1977.

Price, J. L., & Mueller, C. W. A causal model of turnover for nurses, Academy of Management Journal, 1981, 24, 543-565. (a)

Price, J. L., & Mueller, C. W. Professional turnover: The case of nurses. New York: SP Medical and Scientific Books, 1981. (b)

Schneider, J. The "greener grass" phenomenon: Differential effects of a work context alternative on organizational participation and withdrawal intentions. Organizational Behavior and Human Performance, 1976, 16, 308-333.

Smith, P. C., Kendall, L. M., & Hulin, C. L. The measurement of satisfaction in work and retirement. Chicago: Rand McNally, 1969.

Steers, R. M., & Mowday, R. T. Employee turnover and post-decision accommodation processes. In L. L. Cummings and B. M. Staw (Eds.), Research in organizational behavior (Vol 3). Greenwich CT: JAI Press, 1981.

Watson, T. W., & Appel, V. H. The desirability of alternatives: A promising construct for understanding turnover decisions. In Proceedings of the 8th Psychology in the DOD Symposium. U. S. Air Force Academy, CO: USAFA, April 1982.

OFFICER JOB ANALYSIS

Chair: Paul DiTullio

An overview of officer job analysis was provided by representatives of Army, Navy, Marines, Air Force, Coast Guard, and Canadian Occupational Survey organizations.

AD P000830

OFFICER SURVEYS - CANADIAN FORCES

FRED J. HAWRYSH

DIRECTORATE OF MILITARY OCCUPATIONAL STRUCTURES

BACKGROUND

1. Officer surveys have been conducted by the analysis section of the Directorate of Military Occupational Structures (DMOS) since the early seventies. In order to describe how officer surveys are conducted it is first necessary to provide a brief description of some aspects of the Canadian Forces officer system.

2. DMOS exists to support the Military Occupational Structure which is the structural foundation on which personnel management is based. The occupational structure is designed to provide appropriate personnel to meet the roles and objectives of Canadian defence policy. A basic tenet of the occupational structure is that similar task, knowledge and skill requirements are grouped so that personnel in these groups are given orderly and systematic assignments, progressive and economical training and competitive promotion opportunities.

3. Officers of the Canadian Forces are categorized into three broad groups.

   a. General Officers. All Officers of the rank of Brigadier-General and above are known as General Officers. Specifications are not provided for general officers.

   b. General Service Officers. Officers in the general service category of the rank of Colonel and below are trained for and employed in positions relating to their classification and also for general positions which do not call for their specific occupational grouping.

   c. Specialist Officers. Officers in the specialist category in the rank of Colonel and below are normally trained for and employed exclusively in positions relating to their classifications. Some classifications in this group are Medical, Dental, Legal and Chaplain.

4. Within the three broad categories described above there can be any number of classifications. A classification is the basic occupational group into which an officer is assigned. The grouping is based on a requirement to perform related functions embracing similar skills and knowledge associated with the performance of a particular series of duties. Each of these classifications is described by a specification or series of specifications.

5.   Our officers are viewed from two different points of view which in simplistic terms can be described as the Leader/Manager view and the narrow Occupational view. It is this duality of purpose that can make officer surveys difficult. Most officers in a non-specialist category classification (refer to definition of specialist in para 3) spend the majority of their time and effort in the Leader/Manager role while for specialist officers the opposite tends to be true.

6.   Leader/Manager data has been found most difficult to gather by survey, however, it has at the same time been found to be quite uniform in the standard of requirement across all officer classifications. This uniformity is shown in the "Officers General Specification" which describes requirements for all officers regardless of classification. It is from this specification that staff, leadership and management courses are developed. Training which is directed towards occupational requirements does not, of course, exclude the staff, leadership or managerial requirements which are closely related.

HOW SURVEYS WERE CONDUCTED

7.   Officer data has been gathered by survey in three ways:

a. Specific to occupation - (Aircrew Study);

b. Specific to Leadership/Managerial requirements - (Management Study); and

c. Both occupational and Leadership/Managerial requirements in the same survey - (Aerospace Engineer Study).

8.   Officer surveys are conducted on all grades up to and including Colonel. Type of data collected includes:

a. biographical information;

b. job satisfaction;

c. attitudinal information;

d. tasks performed - including time spent and type of involvement in task;

e. knowledge required to perform current duties - including level of knowledge;

f. skills required to perform current duties (eg, pilots aircraft handling skills) - including level of skill; and

g. equipment - including type of involvement with equipment listed.

9. Administrative procedures vary with surveys. Majors and below are usually administered in groups. LCols and Cols receive a short personal briefing and complete the survey on their own. A small percentage of surveys are mailed where circumstances dictate. In this case the surveys are mailed directly to the individual with a personal letter whose purpose is mainly to motivate towards participation in the survey.

SURVEYS COMPLETED

10. The following Officer surveys have been completed:

a. LOGISTICS - All logistics officers MOSs were surveyed as one group. Recommendations made included reversal of a previous decision to combine all logistics officers into a single MOS. This recommendation was not accepted, however, it is anticipated that changes along the lines of the Occupational Analysis recommendations may be made in the next two to three years.

b. AEROSPACE ENGINEERING - Five groups which had previously been separate MOSs were surveyed as a single group. Recommendations included proposing a single MOS with specialties and to include more officer development in both basic and advanced training. These recommendations were accepted and implemented.

c. SECURITY - A single MOS of Security and Intelligence was surveyed in 1974. The main recommendation was that it be split into two MOSs. After much further Branch study the concept was accepted and was put into effect in 1982.

d. AIRCREW - Pilots and Navigators were surveyed together in 1978 (along with non-commissioned aircrew) in the perspective of their participation in the tactical employment of multi-crew aircraft. One result was the modification of the navigator MOS to a single MOS with three sub-classifications. Tactical training for maritime pilots was increased as a result of an identified deficiency. Length of basic training was reduced for Navigators even though the curriculum was expanded. This was accomplished by the removal of "nice to have" segments of training. Officer development training was increased for pilots and navigators as recommended. Navigators were made eligible for more senior officer positions as recommended.

e. MANAGEMENT/TRAINING SURVEY - A sample of Capts to Colonels inclusive participated in this survey in 1975 which was limited to examining managerial requirements. Management courses were updated and upgraded as a result of the survey.

11. Surveys being considered for 1983/1984 timeframe are one of Military Engineering, Marine Engineering, Logistics and Automated Data Processing personnel (both officers and other ranks.

FUTURE DIRECTION

12. In our somewhat chequered OA past in the Canadian Forces we have tended to place greater emphasis on enlisted surveys then on officer studies, primarily because payoffs in terms of any reduction in training and development time are multipled by rather large numbers.

13. During the past few years our methodologies have improved not only in the process of actually doing the studies, but also in the intricies of having the data understood and acted on. Our efforts will be concentrated on refining our current methodology.

AD P000831

OFFICER OCCUPATIONAL SURVEYS IN THE MARINE CORPS

David W. Sutter
Policy/Control Branch
Training Department
Headquarters, U. S. Marine Corps
Washington, D. C. 20380

The Marine Corps has conducted 19 officer occupational surveys since 1971. The problem associated with officer surveys that we all have experienced, i.e., that officers are usually managers and they perform softskill tasks, has precluded the Marine Corps from surveying officers over the entire occupational field spectrum. Officer studies have usually been linked to the technical officer occupational specialties and to special purpose studies.

Officer occupational surveys used the standard task analysis methodology such as was used for enlisted studies - research phase, task list construction, observation and interview, questionnaire construction, administration in person by Marines, CODAP processing, objective analysis, report writing, and subsequent staffing.

Officer grades surveyed included warrant officer through colonel except in special study situations requiring fewer grades. Officers completed a questionnaire booklet that consisted of descriptive information about the incumbent, job related background questions, task statements, job satisfaction (not all studies) and solicited written comments.

A summary of the officer occupational surveys conducted is listed in Table 1.

TABLE 1

OFFICERS OCCUPATIONAL ANALYSIS STUDIES

OccFld/MOS	MOSs Included	Last Admin	Disposition
01 Personnel & Administration	02,07,08,30,60,70 & 80	Sep 72	Appr Jul 75
02 Intelligence	05,02,10,40 & 50	Jan & Feb 75	Appr Apr 77
04 Logistics	02,10,30 & 50	Feb 76	Appr May 79
23 Ammunition and EOD	05 & 10	Jun 74	Appr Aug 75
28 Data/Comm Maintenance	05,10 (Studied w/ OccFld 59 & 593X)	Early 77	Appr Jan 78
30 Supply Admin & Opns	02,10,40 & 60 (Studied w/MOS 9662)	Mar 76	Appr Jan 79
3102 Traffic Mgmt Officer		Feb 76	Appr Mar 78
33 Food Service	02 & 10 (Studied w/ OccFld 41 & 99)	Feb 76	Appr Dec 77
34 Auditing, Fin & Acct	02,06 & 10 (Studied w/MOS 9644)	Feb 76	Appr Dec 78
40 Data Systems	All	Dec 73	Appr Jul 75
41 MC Exchange	30 (Studied w/OccFlds 99XX&33)	Feb 76	Appr Dec 77
59 Electronics Maintenance	05,07,10,20,50 & 70 (Studied w/OccFld 28)	Early 77	Appr Jan 78

SPECIAL STUDIES
OFFICERS

Study	Last Admin	Disposition
01 (Bn Admin)	Jan-Feb 81	Dir, MP
2502 (Comm O)	Jun 77	MOS Specialist
3002 (Ground Supply O)	Jun 81	DC/S I&L
Marine Lieutenants	Apr 80	TBS, MCDEC
Education for Military O	Mar 71	DC/S M HQMC
SEP (Special Ed Prog)	Nov 74	DC/S M HQMC
NAO (Naval Avn Observer)	Mar 82	Dir NAOS, MAG 29

The officer surveys were used for the following purposes:

a. Classification
b. Validation
c. Assignment
d. Training

A recent example of a special purpose training survey is the study conducted for The Basic School, which is provided in enclosure (1). This study provides an objective methodology for curriculum review and design. It could save training dollars by eliminating unnecessary portions of the curriculum; see the table on page 4 of enclosure (1).

All of the studies listed in Table 1 were accomplished under the cognizance of the Manpower Department. On 6 November 1981 the Office of Manpower Utilization (Task Analysis) was incorporated into a reorganized Training Department. The primary mission of the Training Department is to develop policies and programs for the training and education of Regular and Reserve Marine Corps personnel and units. This responsibility includes the formulation, development, and publication of individual and collective training standards for all categories of training conducted in Marine Corps units and institutions.

At the present time, the Training Department has formed an Ad Hoc group to develop and establish the standard operating procedures necessary to complete the Individual Training Standards Manuals (ITSMs). Training standards have been developed for only one occupational field. Training standards need to be developed for 37 occupational fields with a total of 756 military occupational specialties, recruit training, officer acquisition training, professional development education, essential subjects training and related training. Because the Marine Corps uses a manual system that requires a long period of time to produce an ITSM, we are investigating the possibility of automating the ISD process. Our purpose is to develop software that will reduce the time and cost required in the analysis, design and development phases. That is why we are interested in knowing about any software systems that may have already been developed. Then, once we have our process and the first ITSM is completed, the three branches within the Training Department responsible for analysis and ITSM development will begin work on the areas under their cognizance. For example, the Professional Development Education Branch will complete studies of formal officer schools, such as: Command and Staff College, Amphibious Warfare School, and the Basic School.

We have not come to the point of establishing, although it may be soon, a Training Department prioritized survey schedule. In the interim, the Marine Corps has an interest in and intends to monitor the efforts of the other services.

JOB ANALYSIS INVENTORIES IN THE PRIVATE SECTOR

Chair: Ron Page

Papers describing uses of job analysis inventories within private industry were presented. Among the participants were representatives from Control Data Corporation, Honeywell, Inc., American Telephone and Telegraph, and Organizational Research and Development, Inc.

The Use of Job Analysis Information in Assigning Managers to Positions in a Diagnostic Organizational Simulation[1]

Stephen M. Colarelli and Dennis J. Colarelli
Ball Foundation Boulder, Colorado

Within the past few years several corporations began using an organizational simulation called Looking Glass, Inc. (LGI; McCall & Lombardo, Note 1) to analyze the training and development needs of managers. LGI is a six-hour simulation of the top 20 management positions in a medium sized manufacturing organization. The positions range from president to plant manager, and the simulation is organized into three divisions. Each division faces a different external environment: turbulent, placid, and a mixture of the two. The content of the simulation is based on issues and problems faced by managers in actual glass manufacturing organizations (McCall & Lombardo, Note 1). Participants are placed in an office-like setting--complete with a telephone system, mail stations, financial statements, and memos--and they are free to run the organization in any way they please. Looking Glass is a remarkably accurate simulation of a typical "day in the life" of upper level managers. Evidence of the content validity can be found in McCall and Lombardo (Note 1; 1982). In running LGI, the participants produce managerial behavior; this is observed by staff members and used as the basis for individual diagnosis. For a more detailed description of the uses of LGI in training needs analysis, see Kaplan (Note 2).

There are several advantages in using a simulation to assess managers' training and development needs. First, diagnosis is made in an off-the-job environment which is supportive and encourages introspection and open discussion. Second, a trained staff observes participants' behavior. And third, staff observers--unlike colleagues back on the job--are able to view a full range of managerial behavior. However, a critical issue in using LGI for needs analysis is the degree of isomorphism between the behaviors required in a LGI position and those required in the actual position a participant occupies in his employing organization.

If a manager is placed in a LGI position that is substantially different

---

[1] Requests for reprints should be sent to Stephen M. Colarelli, Ball Foundation, 800 Roosevelt Road, Building B, Room 314, Glen Ellyn, Illinois 60137

than his own organizational position, two problems may result. First, the individual may behave less effectively than usual because the unfamiliar demands of the LGI position may require him to spend most of his time simply learning the job. This being the case, the feedback he receives would be inaccurate, not reflective of his true strengths and weaknesses. For example, this would likely occur if an individual with a staff position were assigned to the president's role in LGI. On the other hand, when an individual is placed in a LGI position that is similar to his regular job, the demands of the LGI position are familiar. He will spend little time learning the position and be more likely to exhibit his typical managerial behavior. As a result, the feedback he receives will be more reflective of his actual needs. The purpose of this paper is to describe how job analysis information andan hierarchical assignment algorithm were used to assign managers to positions in the simulation.

Three steps were involved in developing a procedure to match managers to LGI positions. First, a job analysis of the positions in the simulation was carried out. Second, a method for gathering information from participants on their current positions was developed. And third, an assignment procedure was developed to combine the two sets of job analysis information so that each of the 20 participants would be matched to a position in LGI.

<u>Analysis of LGI Positions</u>

A job analysis of the LGI positions was performed using the Position Description Questionnaire (PDQ; Page & Gomez, Note 3). The PDQ is a behaviorally based job analysis instrument developed specifically for managerial jobs. It contains 154 items which measure nine position description dimensions: strategic planning, product/service activities, controlling, monitoring business indicators, supervising, coordinating, customer relations/marketing, external contacts, and consulting. For each item, respondents indicated its combined frequency and importance on a scale from 0 to 4. Four PDQs were independently completed for each LGI position. The first two PDQs were completed by managers from a large manufacturing organization after they participated in the simulation. The final two PDQs for each position were completed by the developers of the simulation and staff members. Final scores were determined by computing the mean score over the four raters for each dimension. This resulted in a profile of nine dimension scores for each position.

A series of 3 x 6 analysis of variance tests (three divisions by six positions) revealed that the positions differed significantly on five of the job dimensions, while significant differences occurred on three dimensions across the three divisions. None of the interaction terms was significant. These results confirmed our belief that different positions in LGI demanded different behaviors. A more thorough description of the job analysis of LGI can be found in Stein (Note 4).

Analysis of Participants' Corporate Positions

The next step was to devise a method for obtaining job analysis information on participants' corporate jobs on the nine PDQ dimensions. Asking managers to complete the PDQ on their corporate jobs was considered but ruled out because of the length of the instrument. Past experience indicated that managers would be reluctant to spend an hour of their time prior to the simulation completing the PDQ. To overcome this problem, a short form of the PDQ was developed (PDQ-SF; Colarelli et al., in press). The final version of the PDQ-SF consisted of 54 items and took about 15 minutes to complete. It closely paralleled the PDQ with respect to factor structure, convergent and discriminant validity, and reliability.

Assigning Participants to LGI Positions

Finally, it was necessary to devise a method to combine the two sets of job analysis information to assign participants to the LGI positions. This problem could be broken down into two tasks. First, the degree of fit between each participant's profile and the optimal profile for each LGI position needed to be determined. Second, an algorithm was needed to assign participants to positions.

We matched participants' scores with the optimal scores for each position by using the absolute deviation or $\underline{1}_1$- metric (Srinivasan & Thompson, 1973).

$$(1) \quad s_{ip} = \sum_j \max (v_{ij} - a_{pj}, 0)$$

where

s_{ip} = the score for participant p at position i

a_{pj} = the score for participant p at dimension j

v_{ij} = the optimal score for position i at dimension j

This metric gives us a condition of "least regret." That is, a participant is penalized to the degree his score falls short of the optimal score, yet he is not given credit when his score is greater than the optimal score. The closer one is to the optimal score on each dimension the better, but scores above the optimal do not count in one's favor. We decided against using a full compensatory model because we wanted to maximize the utility of high dimension scores across all 20 positions (Srinivasan & Thompson, 1973).

With the set of participant-position scores determined, the second task was to assign the participants to the 20 positions. This involved a trade off between individual and organizational utility. To maximize the relevance of individual feedback, we wanted to match each participant with the LGI position that most closely resembled his corporate job. Yet it was also important to be concerned about the effectiveness of the total

organization. If the organization performed ineffectively, the quality of the experience and feedback for all participants would be lessened.

The standard assignment problem solution that optimizes organizational efficiency is given below (Charnes, Cooper, Niehaus, & Stedry, 1969; Srinivasan & Thompson, 1973):

$$(2)\quad \text{Min} \sum_i \sum_j s_{ip}\, x_{ip}$$

subject to the constraints

$$\sum_i x_{ip} = 1 \text{ for all } p$$

$$\sum_p x_{ip} = 1 \text{ for all } i$$

$$x_{ip} = 0 \text{ or } 1 \text{ for all } i \text{ and } p$$

where

s_{ip} = the score for participant p at position i

x_{ip} = amount of position i assigned to participant p

However, with this solution some participants who are best matched for a given position may be assigned elsewhere so that the total system is optimized, thus jeopardizing individual utility. On the other hand, there are situations where maximizing individual utility (i.e., in this case, maximizing the fit between persons and positions) may jeopardize organizational utility. Consider the example presented in Figure 1. If one

Figure 1

Individual Utility Scores

Candidate	Jobs		
	President	Director	Plant Manager
A	20	0	100
B	40	10	85
C	75	95	50

Note: 0 equals a perfect match between corporate position PDQ scores and LGI position.

were concerned with maximizing individual utility, candidate A would be assigned to the director's role, leaving candidate B to the president's role, and candidate C to the plant manager's role. This procedure placed the second most qualified person in the president's role. It is, however, reasonable to assume that organizational effectiveness is hampered when the most qualified candidate for the president's role is placed elsewhere. To deal with the problems of the differential contributions of the positions to organizational effectiveness and the importance of a good person-position match for training needs diagnosis, we used the following hierarchial assignment procedure:

$$(3) \quad \text{Min} \sum_p s_{ip} x_{ip}$$

where $i = 1, 2, \ldots 20$

subject to the constraints in (2)

All positions were ranked according to their importance in contributing to organizational effectiveness. The president was ranked first, vice president second, and so on through plant manager. Positions were assigned sequentially, starting with the president and filling it with the best matched individual from all 20 participants. That participant was then removed from the list of available participants and the next most important position was filled. This continued until all the positions were filled. This procedure allowed a compromise to be reached between individual and organizational utility.

Discussion

In this paper we described how job analysis information and an hierarchical assignment algorithm were used to assign participants to positions in an organizational simulation that is being used by several corporations as a tool for analyzing managerial training needs.

A unique feature was that job analysis information was used both for determining job content and in prediction. The job analysis of LGI positions determined the job content, and the information from each participant's analysis of his corporate job served as the predictor. Two key differences exist between this and traditional assessment and selection methods. First, we used dimensional content of individuals' current jobs as predictors, not ability measures. Second, we were not trying to differentiate individuals on qualities in order to predict job performance over the long run. Rather, our goal was to differentiate people according to the demands of their current jobs in order to maximize performance in a temporary system over the short run. That is, we wanted to place people in positions so as to minimize the new learning that needed to occur for effective performance, not place people based on their potential to learn the demands of the positions.

One must note, however, that our assumption about the validity of the procedure remains untested. We have no empirical data on the effectiveness of assigning people to LGI jobs most similar to their own. Thus we cannot be sure that participants assigned to LGI by the procedure described here actually performed more effectively (and hence received more accurate feedback) than, say, if they had been randomly assigned. At this point, the soundness of the procedure rests upon its content validity.

Another interesting aspect was the hierarchical assignment algorithm. Most multi-attribute assignment algorithms are concerned with maximizing organizational utility (Charnes, et al., 1969). The assignment model described here was constructed to reach a compromise between both organizational and individual utility. Given the goals for the simulation, organizational and individual utility were not independent, and thus both had to be taken into account. However, while our model was suitable, it was not necessarily optimal. Comparisons are needed between the present and other assignment procedures to determine the most effective model for maximizing both utility parameters.

Reference Notes

1. McCall, M.M., & Lombardo, M.M. Looking Glass, Inc.: An organizational simulation. Technical Report No. 12. Greensboro, North Carolina: Center for Creative Leadership, 1978.
2. Kaplan, R.E. Management simulation: A new tool for self-directed needs analysis. Paper presented at the meeting of the Academy of Management, New York, New York, 1982.
3. Page, R.C., & Gomez, L.R. The development and application of job evaluation and staff systems using the Position Description Questionnaire. Personnel Research Report No. 162-79. Minneapolis: Control Data Corporation, 1979.
4. Stein, R.T. A job analysis of Looking Glass, Inc.: A study of content validity. Unpublished manuscript, Center for Creative Leadership, 1980.

References

Charnes, A., Cooper, W.W., Niehaus, R.J., & Stedry, A. Static and dynamic assignment models with multiple objectives, and some remarks on organization design. Management Science, 1969, 15, 365 - 373.

Colarelli, S.M., Stumpf, S.A., & Wall, S.J. Cross-validation of the Position Description Questionnaire. Educational and Psychological Measurement, in press.

McCall, M.M., & Lombardo, M.M. Using simulation for leadership and management research: Through the looking glass. Management Science, 1982, 28, 553-549.

Strinivasan, V., & Thompson, G.L. Al..nate formulations for static multi-attribute assignment models. Management Science, 1973, 20, 154-158.

AD P000833

The Use of Job Analysis Inventories for Developing Improved Selection Systems

Gail Drauden, Sr. Personnel Research Specialist, Honeywell

The use of job analysis material is basic to the choice, construction or validation of materials used in selecting people in entry or promotional situations.

Task/ksa inventory analysis is better suited to some of the tasks facing the selection expert than for others. For most of the work: examining job design, classification of jobs, constructing work samples, and interview and other materials, the task/ksa inventory is the single most useful tool.

First of all, in designing a selection project, the analyst must determine which jobs belong together. We usually do this by making a rational determination of which job titles belong in the family, then using the task inventory to examine the structure of the jobs within the family. While there is some controversy these days as to whether it is necessary to break jobs into their components and use perhaps somewhat different batteries for job groups of different component mixes, it is always the case that the analyst wants to know this, if only to be able to tease out reasons post hoc for anomalous results. The first step, then is to determine the homogeneity of the job family, and decide whether to treat the study as one of a single or multiple groups. Related to this, and at this phase of the work, we also bring the preliminary results of the job analysis back to the client manager's to review. Is this the way they want the job to be done? Task/ksa inventory printouts are a fast way for the managers to see how the job is presently being done and to work at restructuring the positions before a massive testing study is carried out on jobs that should be redefined.

We find this classification step most helpful in large studies, spanning many divisions and job titles. The interdivisional studies of Production Control, Factory Supervisor, and Sales Representative are examples of such large projects. In each of these, we needed to examine the similarity of work across organizational units, job titles, and geographic area.

A second place task/ksa inventories are useful are in the construction of work sample tests. I find this to be true especially in operator and technician work. For assembly and technical areas, there is always an engineer or some other expert who can help construct the proper materials once the important knowledges, skills, and task compentencies have been identified. In white collar jobs, such as management and sales, the work domain has not been as clearly identified, and the test constructor finds that he or she must do much of the test construction. White collar jobs require the analyst to build the test and to obtain additional information through interviews and critical incident gathering. With operator, or technical jobs, the results of the task/ksa inventory can be turned over to some job expert, and with coaching on psychometrics and test construction issues, that person or group can build the test themselves. I used this approach in one of our factories which does assembly of large circuit boards. The task/ksa inventory results indicated the things that entry operators usually did with high frequency. An industrial engineer in the plant built these tasks and skills into a work sample. We are conducting a predictive study on this test (and some ability and dexterity tests we are administering along with it), but on preliminary analysis it seems to be the best predictor. A similar approach is being taken in building an entry test for one of our high technology microscope assembly clean room areas. A task analysis I did there last year identified the major tasks, skills, and personality requirements. Job experts will be building similations of some of the tasks, and I have chosen ability and personality tests on the basis of the inventory output.

As a practical matter, these inventory results can be used like specifications. Having this clear list of what is needed cuts down on the discussion time required to get a group of experts moving toward a finished product. It focusses discussion and speeds the work.

A third area in which task/ksa inventories are useful in selection work is as a basis for test choice, for picking the best from the standard published stock. I usually come back from the round of interviews from which the inventories are constructed, with some pretty solid ideas about what abilities are required on the job. I find the process of looking at the highest frequency or importance tasks and ksas, and looking at the factor structures that emerge, suggest additional areas and changes in emphases on the areas already identified. The inventory process enriches, organizes, and refines the judgments made on the basis of the interviews alone.

Others are discussing inventories in relation to performance appraisal and evaluation. I will just mention that our procedure is to use the dimensions resulting from the task/ksa inventory, collect critical incidents on them and contruct BOS scales on them, as the criteria for most of our validation work.

We also rely heavily on task/ksa inventories for constructing the support materials used in a selection system. By this I mean those products which are usually content validated: jobs previews, training and experience ratings, and interview questions.

One approach which we have used as a sort of stop-gap measure to provide reasonable selection supports until more structured testing could be done, is a combination job preview and training & experience booklet. The booklet describes the various ksa's required on the job, one per page. For each ksa, the tasks associated with it are described. Then the applicant is asked when he or she had ever done anything of this type. For example: "Ability to teach or train:: The sales representative will generally organize and participate in teaching 3-5 large seminars (20 people for 2-3 days) during the year. The representative, depending on the audience, will train in sales techniques, technical product information...." Such booklets can be constructed almost by the numbers from the results of task/ksa inventories.

A related product that we provide our divisions is the structured interview form. These interview forms usually have one part built from the task/ksa inventory which list the ksa factor, then indicate various tasks on the job in which the ksa factor is used. The interviewer is told to ask about skills and experience in these areas. We also include job demands on our inventories, such things as required travel, noise, dirt, outdoor work, long hours, interruptions at home, etc. The interviewer is told to ask the applicant how he/she feels about these jobs demands. The second part of the interview comes from critical incidents, and would ask the applicant what he/she would do in given situations.

In addition to the actual construction, choice, and validation of selection instruments, there are some administrative and organizational reasons why task/ksa inventories are useful.

One reason is consistency. We use the dimensions resulting from the task/ksa inventories as a basis for an integrated approach to personnel programs. This enables us to use the same dimensions for all of our selection materials, our performance appraisal materials, our training needs analysis, and so forth. Because the inventories are carefully constructed from interviews and tailor made for the job families, and because a large sample of incumbents and managers complete the forms, we are confident that the results are a solid base on which to build out programs.

A second reason is speed. For large programs especially, the initial investment in time in doing the job analysis survey is repaid by the speed with which some of the later products can be delivered. We are all aware of the amount and the richness of the information that job analysis inventories can gather. I want to emphasize the order and clarity of the results. This has allowed us to put together as many as 70 interview booklets in two weeks, from the printouts, in rather mechanical fashion, and yet produce what I subjectively feel were good forms.

A third reason, and perhaps one of the most important, for using task/ksa inventories is the higher probability of acceptance for the resulting selection materials. Employees know that the survey itself came from employee interviews and observations. The results of the survey also come directly from the responses of people doing the job. Two outcomes seem to flow from this.

One outcome is that the basis of whatever personnel program is instituted is clear and visible. When we helped implement a system of promotional interviews based on a job analysis inventory completed throughout a factory, there was no confusion among those going through the promotional process as to where the questions came from. Material on orientation booklets, structured interviews, etc. can be traced directly back to the inventory and the items ranked as most frequently performed or most important. A typical employee response is, "I remember that item from the survey."

A second outcome is that employee ownership of the programs is increased. Because people in the job are included from the very beginning, in the job description and program planning stage, they perceive personnel programs based on a task/ksa inventory as something they have done, have had a part in, helped develop. The job inventory booklets which reach everyone or nearly everyone in the job, give these members of the job class an opportunity to impact on the content of the program. During the interview and survey process, we explain that the purpose is to obtain information from which to design some specific program. When the program (structured interview, performance appraisal form, etc.) is implemented, any objections to it can be dealt with by demonstrating that the basis of the program was the information provided by the employees and managers themselves.

I have tried to demonstrate the usefulness of task/ksa inventories in the areas of job grouping, work design evaluation before a study is done, construction of work sample tests (especially in factory operator and technician jobs), selection of ability or personality tests, and the construction of training and experience and interview questions.

Finally, I would like to emphasize once more that the power task/ksa inventories have for involving everyone in the development of personnel programs and establishing ownership for the programs among the people who will be affected by them is a little mentioned but very important reason to use them.

JOB ANALYSIS AND TASK ORIENTED RATINGS

Sidney Gael[1]
American Telephone and Telegraph Company

The present paper describes two recent task oriented rating (TOR) studies conducted in the Bell System in which TOR forms based on job inventory results were developed and utilized. The study objectives were to (1) examine the psychometric properties of TORs, and (2) determine whether supervisors can use TORs to evaluate task by task effectiveness. Several job analytic based TOR and TOR-like approaches have been studied for many jobs with generally favorable results. Rosinger, et. al. (1982), for instance, found little rater error, high interrater reliabiltiy (r=.90), and significant concurrent validity. TORs are an efficient way to evaluate job performance because the rating forms are easily developed once job tasks have been analyzed, performance can be rated in about five to ten minutes per employee, results can be used for employee development and appraisal feedback sessions, and TORs have high face validity because important job tasks are covered.

Work Performance Survey System

Before discussing the two Bell System TOR studies mentioned above, I would like to describe the Work Performance Survey System (WPSS), the job inventory approach we have developed at AT&T. WPSS is a computer assisted job analysis approach that relies on job inventory questionnaires to obtain detailed data about job tasks, functions, and incumbents. Any source of job information from which job tasks can be derived is fair game. Generally, though, the bulk of the task statements contained in WPSS questionnaires is derived through a combination of interviews with job incumbents and supervisors and through content analyses of written materials, such as job descriptions, training materials, maintenance manuals, and company practices. After a WPSS questionnaire has been prepared, trialed, finalized, and printed, copies are distributed with detailed distribution instructions to field coordinators who are responsible for getting the questionnaires to appropriate respondents, tracking project progress, and assuring that completed questionnaires are returned for computerization and analysis.

WPSS questionnaires usually contain two questions about each task statement, for instance, a question about task significance and one about task time. The number of questions is limited to avoid overburdening respondents and operations; the guideline is two hours. The answers requested are ratings on a 0-7 scale, where zero indicates that the incumbent does not perform the task and 1-7 represent differing degrees (low to high) of an attribute. The zero response, in essence, answers the implied question, "Do you perform the task?" in place of a separate question addressing task occurrence.

There are two types of WPSS reports -- statistical summaries and crosstabulations -- but a variety of computer printouts can be obtained within each type. Figure 1 shows a sample taken from a task significance by company report. The column headings in Figure 1 represent a total sample and telephone company subsamples. Similar reports can be generated for job title subsamples instead of company subsamples. Separate reports must be generated for each task attribute question contained in a WPSS questionnaire, i.e., task attributes cannot be used as column headings. The number of total sample and subsample respondents appears immediately beneath the column

[1] The views expressed here are solely those of the author's and not necessarily those of his employer.

headings. Each cell in the report contains a mean response for a task statement, a standard deviation, and the proportion of the specific sample that contributed to the cell statistics. Reports summarizing the relative percent time spent on tasks and functions can also be generated, providing that a time spent question is included in the questionnaire. WPSS software, by the way, is available to the general public for a fee under a license agreement with AT&T.

SURVEY: 1000

STATISTICS FOR SIGNIFICANCE

SIGNIFICANCE		COMPANY TOTAL	C&P	ILLINOIS	INDIANA	MICHIGAN
	N=	113	5	6	2	5
ROW TOTAL	PROP	1.00	1.00	1.00	1.00	1.00
	MEAN	3.259	3.672	2.649	3.020	3.043
	STD	1.905	1.961	1.671	1.766	1.576
1. ANALYZE OFFICE	PROP	0.97	1.00	0.83	1.00	1.00
COSTS/EXPENSES FOR	MEAN	3.673	3.400	2.800	5.000	4.400
BUDGETING PURPOSE	STD	1.714	0.894	1.789	1.414	2.302
2. ASSIGN DEVELOP-	PROP	0.92	0.80	0.83	1.00	[illegible]
MENT OF LONG RANGE	MEAN	4.048	4.000	4.000	6.500	3.000
FORCE/TRNG REQRMNT	STD	1.787	1.155	1.414	0.707	2.000
[illegible] DETERMINE	PROP	0.90	1.00	0.83	1.00	0.80
P[illegible]NNEL NEEDS FOR	MEAN	3.578	5.000	2.800	3.500	3.000
A N[illegible] PROJECT	STD	1.714	2.000	2.049	2.121	1.414
4. DETERMINE	PROP	0.92	0.60	0.33	1.00	0.80
WORK FLOW &	MEAN	3.375	4.333	3.000	2.500	3.250
OFFICE DESIGN LAYO	STD	1.769	1.528	2.121	2.121	1.258
5. DEVELOP	PROP	0.79	0.80	1.00	1.00	0.40
ABSENCE & TARDINESS	MEAN	3.382	3.75[illegible]	3.000	4.500	2.000
POLICY	STD	1.655	0.5[illegible]	2.000	2.121	1.414
6. DEVELOP	PROP	0.97	1.00	1.00	1.00	0.80
ANNUAL FORCE PROGRAM	MEAN	4.691	5.600	3.500	6.500	4.250
	STD	1.801	1.342	1.04[illegible]	0.707	2.500
7. DEVELOP	PROP	0.83	0.60	1.00	1.00	0.80
DISTRICT BUDGET &	MEAN	4.309	6.333	2.833	3.500	4.250
EXPENSE PROGRAM	STD	1.878	1.155	1.941	[illegible].536	2.062

Figure 1. Sample from a WPSS statistical summary report.

Cell statistics in conjunction with specific selection criteria will help identify tasks that should be considered further, for instance, for inclusion in a TOR form. Selection criteria, though arbitrary, should be toward the high end of the significance or importance scale. One way to proceed, as per the present studies, is to single out tasks that attain at least a five average rating on a seven point significance scale scale and then concentrate only on those tasks that are performed by at least 50 percent of the job incumbents surveyed.

TOR Studies for Telephone Company Jobs

One of the TOR studies pertains to technicians who maintain switching equipment, and the other to service representatives (SRs) who either sell equipment and services or handle billing and collection work. After the technician TOR data were analyzed, we realized that another study was needed to compensate for the fact that technicians are on their own for long time periods and supervisors do not have the opportunity to observe them perform

many important tasks. SRs and their supervisors, on the other hand, work closely together, and the supervisors can easily observe the full range of SR tasks.

Both studies were conducted in essentially the same way. TOR forms and an orientation for supervisors were developed, trialed, and modified in accordance with tryout results; small groups of supervisors per job met at several locations throughout the country, the orientation was presented, and the supervisors completed TOR forms for each of their subordinates plus a few other employees not under their direct supervision whose work they they felt they could rate. Supervisors also rank ordered subordinates on the basis of overall effectiveness, and indicated the importance of each SR function performed in their particular operation by spreading 100 points across the functions. Technician supervisors rank ordered their subordinates again one month later. Inasmuch as the two sets of rank orders were very highly correlated ($r=.82$), the procedure was not repeated with SR supervisors. SR supervisors were asked to provide indices of individual SR output tracked during day to day operations (performance records are not maintained systematically for individual technicians) and to answer a few questions about the acceptability of the TOR approach.

The technician TOR form was composed of 36 task statements, four under each of nine functions, the sales SR form was composed of 40 task statements under seven functions, and the billing SR form was composed of 42 task statements under six functions. A seven category rating scale, as recommended by Siegel, Federman, and Wesland (1980), was used to rate task performance effectiveness, and the words, slightly, somewhat, rather, quite, decidedly, very, and extremely, taken from a perceived intensity scale (Bass, Cascio, and O'Connor, 1973) were used to label each effectiveness category. Two additional categories were provided so that supervisors could identify tasks that are not part of an incumbent's job or that they had not observed being performed.

The purpose of the rater orientation was primarily to minimize rater errors commonly found in rating studies. The orientations, presented immediately prior to the rating session, were standardized in terms of content, presentation sequence, visual aids, and time devoted to each topic. Each orientation included the purpose of the study, including its research perspective, an assurance of confidentiality, a review of the TOR form, and a description of rating pitfalls, e.g., halo, leniency, and central tendency, and how they might be avoided. The content of the orientation for SR supervisors was expanded by including a discussion of a systematic thought process they might follow when evaluating task performance, more detailed discussions of the TOR rating process and rating pitfalls, plus rating exercises that the supervisors completed and discussed before rating their subordinates. The expanded rater orientation was intended to increase rater reliability over that obtained in the technician study, which is in the range typically obtained in job performance rating studies.

TOR Study Results

Supervisors completed TORs for employees in the three jobs studied as follows:

	Sup.	Tech.	Sup.	SR-Billing	Sup.	SR-Sales
Primary	23	138	33	259	30	220
Secondary	20	52	13	33	13	31

As can be seen above, supervisors were sufficiently familiar with only a few employees not under their direct supervison to rate (secondary) their task by task performance; only 19 percent of the 617 subordinates could be rated by a second supervisor. As expected, the Not Observed and Not Performed categories were used much less frequently, on the average, for the sales and billing SR jobs (10.5 and 7.3 percent, respectively) then for the technician job (32.9 percent).

Two methods were used to score TORs. For both methods, task ratings were first averaged by function per employee, and function averages, in turn, were averaged to obtain total TOR values. For one of the two methods though, function averages were multiplied by weights derived from the assignments of 100 points across the functions before they were averaged to obtain total TOR values.

TOR Discriminating Power

Regardless of job classification or scoring method, TOR values obtained support the discriminating power of TORs. The distributions, however, are negatively skewed, suggesting a tendency toward leniency. Individual task rating distributions indicated that the lower rating categories were used infrequently. If the task rating distributions represent a tendency toward leniency, it is not due to just a few tasks, but pervades the task ratings. Another explanation for the task rating distributions is that employees included in the study, for the most part, warrant ratings above the mid-scale value. Function raw score averages ranged from 4.8 to 5.6 for technicians and 5.1 to 5.9 for SRs, again suggesting a tendency toward leniency, but the magnitudes of the standard deviations (about 1.3, on the average, for technicians and 1.1 for SRs) indicate that the distributions around function averages are adequate.

Other views of the discriminating power of TORs were obtained by correlating individual task ratings with total TOR values for the SR jobs and through analyses of variance. The correlations obtained are moderate to high for both jobs, suggesting that task ratings are measuring the same underlying dimension(s) as the total value, and that they discriminate those with high from those with low total TOR values. On the other hand, the correlations may be regarded as indicative of halo. Ratee main effects in analyses of variance, except for one study site, are highly significant, supporting the discriminating power of the TORs. The ratee effect for raw scores accounted for about 45 percent of the variance on the average across SR study sites, whereas the ratee effect for weighted scores accounted for 13 percent of the variance.

Rater Reliability and Validity

Rater reliability was determined by correlating two independent ratings obtained for each employee. The average correlation for pairs of technician supervisors is .46. As mentioned previously, reliability coefficients obtained for rank orderings of technicians is .82. TOR reliability coefficients for billing and sales SRs are .56 and .16, respectively. The low reliability obtained for sales SRs is due mainly to a few pairs of ratings -- without all four pairs of ratings obtained at one location and the single pair obtained at another location, the correlation coefficient for the remaining 26 pairs of ratings is .36, still quite low. TOR validity was determined by correlating TOR values with standardized ranks and several performance indices tracked daily for individual SRs. Correlations of TOR values with ranks are .71, .56, and .69 for the technician, and billing and sales SR jobs, respectively (highly significant in each case). Five job

performance indices were obtained for 54 sales SRs, while only one index tracked for only 25 billing SRs, the percent of incoming calls handled (PCH), was suitable for analysis. The correlations between the performance indices and the raw score and weighted TOR values are as follows:

	Raw Score	Weighted
Sales Volume	.31	.26
Percent Items Sold	.26	.31
Gift Certificate Sales	.47	.51
Service Order Accuracy	.31	.33
Phone Center Store Referrals	.24	.29

The above correlations are statistically significant at least at the .05 level, except for the correlation between TOR raw score values and Phone Center Store Referrals. The relationship between TOR values and the one performance indicator obtained for billing SRs, PCH, was not statistically significant.

Sources of Rater Error

Rating approaches should seek to minimize certain rater tendencies, such as tendencies to focus on global impressions rather than distinguish among different aspects of performance (halo), to be too lenient or severe, or to concentrate ratings at the midpoint of the rating scale (central tendency). A few ways by which the presence of rater error was examined have already been mentioned. Additional views of rater error, in accordance with operational definitions of rater error found in the literature (Saal, Downey, and Lahey, 1980) are discussed below.

Halo. Function intercorrelations, correlations between task ratings and total TOR values (previously discussed), principal components analyses of task ratings, and rater by ratee interactions in analyses of variance indicated that halo was present in varying degrees. Many function intercorrelations, for instance, range between .40 and .70 suggesting the presence of halo, but there are also logical grounds for the relationships between functions. The proportion of variance accounted for by the rater by ratee interaction in the technician study is less than four percent for the analysis of raw scores and about one percent for the weighted scores, whereas for the billing and sales SRs, the proportions are 15 and 5 percent and 17 and 3 percent, respectively.

Leniency/Severity, and Central Tendency. Distributions and statistics for tasks, functions, and total TOR values and rater main effects in analyses of variance were used to determine the presence of leniency (or severity) in the ratings. As mentioned previously, the function averages and the negatively skewed distributions indicated that leniency might be present. By the same token, central tendency was ruled out. Statistically significant rater main effects also signified the presence leniency, but the proportions of variance accounted for by the effect are neligible.

Degree of Acceptance

The utility of any performance evaluation approach is dependent upon its acceptance by those who use it. In response to questions designed to obtain impressions of TORs on five point scales, supervisors mainly used the top

three categories to rate TOR fairness (96.8%), objectivity (90.5%), ease of use (73%), acceptability to employees (79.3%), and favorability compared to the present evaluation procedure (78.7%). A few procedural modifications were suggested, for example, expand the task list to include more job activities.

Conclusions

It appears that supervisors can use TORs adequately, and, according to their questionnaire responses, would like to use them in on-going operations. The data suggest that rater errors commonly associated with ratings are in evidence, but the effects appear to be small and should not interfere with the application of the TOR approach. In any case, the rater errors found give employees the benefit of the doubt. The largest proportion of variance accounted for is associated with ratee effects, where it should be. The extended training for supervisors introduced in the SR study did not help increase the reliability over that obtained in the technician study. As Saal, Downey, and Lahey (1980) point out, however, a number of researchers have expressed reservations about the usefulness of interrater reliability or agreement as a criterion of rating quality. Borman, for instance, found that reducing rater error through training produced lower interrater reliability but more accurate performance profiles. A significant problem confronting reliability analyses conducted in industry is finding two or more supervisors who are in a position to rate the same employee. Certaiinly, the validity analyses for the present studies are highly supportive of the TOR approach. Inasmuch as weighted score values affected distributions and proportions of rater error variance more favorably than raw score values, and since they have greater face validity, they will continue to be used.

In view of the study results, a much broader trial of TORs will be initiated. Since the extended training did not produce anticipated results, the next time around, TORs will be introduced in some groups without any special training at all. Perhaps, special training for so simple an instrument may not be buying anything. A sought after ingredient in the anticipated trial will be jobs in which individual employee performance indices are systematically tracked. Many administrative procedures for actual TOR implementation remain to be worked out, for example, developing a weighted scoring method that can be used across locations, setting standards and developing a total evaluation approach, and then, of course, preparing methods for using TOR results in appraisal.

References

Bass, B.M., Casio, W.F., and O'Connor, E.J. Magnitude estimations of expressions of frequency and amount, Journal of Applied Psychology, 1974, 59, 313-320..

Rosinger, G., Myers, L.B., Levy, G.W., Loar, M., Mohrman, S.A., and Stock, J.R. Development of a Behavioraly based performance appraisal system, Personnel Psychology, 1982, 35, 75-88.

Saal, F.E., Downey, R.G., and Lahey, M.A. Rating the ratings: Assessing the psychometric quality of rating data. Psychological Bulletin, 1980, 88, 413-428.

Siegel, A.I., Federman, P.J., and Wesland, E. Perceptual/psychomotor requirements basic to 35 Air Force Specialties. AFHRL-TR-30-26, Brooks Air Force Base, Texas.

THE USE OF JOB ANALYSIS INVENTORIES IN JOB EVALUATION

Ronald C. Page, Ph.D. Jeffrey J. McHenry

Control Data Corporation

Control Data Corporation (CDC) is a rather young company. It began in 1957 as a three-person R&D firm in the fledgling computer industry. Many have characterized the computer industry as a growth industry, and Control Data is a good example. By 1963, the company had 3,000 employees, and by 1969, 49,000 employees. In the last decade, the company has matured and diversified from a manufacturer of computer mainframes and peripherals to a service-oriented company providing financial services, data services, education, and health care; and it has undertaken large-scale projects in urban and rural development. Today Control Data employees work in 47 countries.

~~This~~ rapid growth and diversification has presented a number of complex problems to Control Data's personnel function, particularly in compensation. There have been two primary reasons for this: 1) continued technological advances have led to rapid changes in job content that could not be tracked with traditional job analysis methods; and 2) the geographic dispersion of the work force has hindered ensuring equity in compensation.

These factors contributed to the realization that the corporation needed an improved means of evaluating the worth of jobs and compensating employees appropriately. In fact, we realized that our fundamental need was to find out just <u>what</u> work our employees were performing and what constituted the <u>differences</u> between jobs and pay grades. We concluded that we needed an improved means of gathering job content information so that evaluation criteria could be firmly based upon identifiable differences in job content.

After a careful analysis of our problems and needs, we decided on a structured questionnaire approach to job analysis in which questionnaires are tailored to specific job families. The system was to have an integrative role: the job content information would provide inputs for various other personnel functions, including staffing, performance appraisal, training, EEO, career development, and compensation. However, because the corporate sponsor for this undertaking was the Compensation Department, our foremost research applications to date have been in that area.

Initial R&D for a questionnaire-based, computer-scored system for describing and evaluating managerial jobs began in 1974, and the results of this research are documented by Tornow and Pinto (1976) and Gomez, Page, and Tornow (1979, 1982). In developing this management job analysis and job evaluation system, questionnaire design and software development were done in-house. For our non-management jobs, it was decided to adopt the task inventory methodology developed by the Air Force and to acquire CODAP for analyzing the job analysis data. After a period of negotiation in late 1978, we were able to acquire a copy of CODAP from the University of Texas, which had converted the 1974 Univac

version of CODAP to a CDC-compatible version. In 1979, we began the development and implementation of a CODAP-based job analysis and evaluation system for the corporation's 54,000 non-management positions.

After some of our initial publications on our research, we had a number of requests from other companies for assistance in developing similar questionnaire-based systems or for assistance in computer analyzing data that had been collected. Initially, we referred these contacts to others, such as the Air Force, but in 1980 a management consulting susbsidiary was formed, Control Data Business Advisors, Inc. (CDBAI). Although the majority of CDBAI services are oriented toward small businesses, job analysis services have been included with CDBAI's offerings. As a result, we have two organizations within Control Data that are providing job analysis research development and delivery--one research team serving the internal organization, and a second serving external organizations.

Now that we have given you some background on our entry into questionnaire-based job analysis research, we would like to give you an overview of some of the differences between our methodology and that of the Armed Services, and we would also like to describe some of the enhancements and modifications we have made to CODAP. The major areas in which we have worked include 1) making the CODAP system and reports more user-friendly, and 2) manipulating the data so that we can better identify a hierarchy of jobs and determine job value.

Making CODAP user-friendly. Even though the original CODAP programs proved a tremendous aid to our ongoing job analysis and job evaluation efforts, we soon encountered a number of situations where we felt that it would be to our benefit to make the CODAP system more user-friendly. For example, we quickly noticed that it was difficult to track all of the control card and generated files required to perform a CODAP analysis for a given data base. Our solution was to develop a series of interactive front-end programs for CODAP. These front-end programs automatically track all of the CODAP files and, through a few simple prompts of the user, generate all of the job control cards needed to execute the CODAP program and route output to the appropriate output device.

Last year, we completed the development of an automated Task Inventory Management System, TIMS. TIMS is a user-friendly program designed for use by personnel analysts who have little or no computer experience. TIMS assists in the inventory construction process by permitting the easy tracking, sorting, editing, and printing of task inventories during the iterative development process. The TIMS editor allows inventory developers to search previous task inventories for key phrases, thereby assiting in the development of a preliminary inventory. Users can then more effectively work with committees of employees in revising and updating statements. Rather than retyping all of the task statements during each iteration of the inventory development process, revisions are easily entered into the terminal by the personnel analyst, and a new sorting of tasks by duties is quickly printed. The new and revised task statements then become part of the

expanding TIMS data base. TIMS can also process a finalized task inventory and create control cards for CODAP's INPSTD program. As anyone who has created INPSTD control cards knows, these routines can save an enormous amount of data entry time.

User-friendly report formats. Over the past few years, a substantial effort has also been invested in making the CODAP report formats more user-friendly. Our first task was to display output on 8.5 x 11 inch paper to aid storage and use by management. Subsequently, we began reformatting the reports themselves. For example, the PRIJOB program is used to compare the percent time spent performing a given task by various user-specified groups within a data base. PRIJOB output can be useful in determining why groups do or do not form clusters. Unfortunately, PRIJOB output is often extremely bulky. To reduce this bulk, we added two cutoff options to the PRIJOB program. The user could *a priori* specify that the percent time spent by a group on a given task or duty would only be printed if the group spent at least a certain minimum percent of time on that task and/or a certain minimum percent of group members performed that task. Judicious specification of these options reduces the PRIJOB output to a manageable amount and at the same time permits a quick assessment of the critical differences between jobs.

We have also undertaken efforts to upgrade the way we display our information for presentation to management. While the information included in CODAP output is extremely valuable to personnel professionals, we have found that it is not very user-friendly for managers or job incumbents. Therefore, we are investing considerable energy in improving the format of our job analysis output, using computer graphics whenever possible. For example, for our management sytem, we have created a Factor Profile that uses high-speed computer graphics to produce a report summarizing an incumbent's overall standing with respect to sets of job evaluation and job description factors. The Job Comparison Profile is a multi-color pen plot graph that performs much the same function as PRIJOB. The profile shows the percent time spent on all duty areas by two or more groups of job incumbents. These reports contain basically the same information presented in CODAP reports, but the information is now presented in a format that facilitates its understanding by management.

Computational changes. In addition to the changes designed to make the CODAP system more user-friendly, several computational options have been added to CODAP. One change was in the scaling of the time spent responses. We noticed that our individual job descriptions from JOBIND and our group job descriptions from JOBSPC indicated that our employees were all performing an extensive range of tasks, and spending approximately equal amounts of time performing each task. Common sense, however, led us to believe that our job holders had to be spending differentially larger proportions of time on certain tasks. After analyzing a number of job descriptions and relating them to what we knew about these jobs, we realized that our scaling was leading to these bland job descriptions and that we needed a geometrically progressive scale. After contacting the Air Force, we decided to investigate three scaling alternatives: 1.5^x, 1.75^x, and 2.0^x,

where x ranges from one to nine and represents the individual's time spent response. Our results, as judged by subject matter experts, revealed that the 1.75^x and 2.0^x scales yielded more accurate results than either the 1-9 scale or the 1.5^x scale. Our results indicated no clear superiority for either the 1.75^x or the 2.0^x scale, and we have used both in our research. However, we have generally favored the 2.0^x scale because the scale anchors are then round numbers (i.e., 2, 4, 8, 16, etc.).

The second change we made was to transform the data used in the calculation of the similarity values entered into the hierarchical cluster analysis. CODAP's clustering algorithm uses the overlap in percent time spent as the similarity index. We found that using this similarity index yielded groups that were functionally similar. However, the job structures of the clustered groups were often quite dissimilar. This was a major problem for our compensation staff, since we wanted to develop our compensation system from job clusters that reflected job structure. For example, for the 1,892 software employees whom we surveyed with a 445-item questionnaire, we found a rather diverse cluster of incumbents who were grouped together because they shared a number of data entry tasks. This cluster included both keypunch operators and programmer analysts who worked at remote sites and had to handle all of their own data entry. We realized that our rational assessment of similarity was not being captured by the CODAP clustering algorithm. Our rational assessment of similarity was a function of both percent overlap in job behaviors and the organizational value of those shared behaviors. The CODAP clustering procedure, however, was assuming that all tasks should be weighted equally in the assessment of job similarity. We therefore decided to estimate the value of each task to the organization through a process we call task valuing. The time spent index was then weighted by multiplying the time spent times the task value. The resulting weighted time spent index is used to compute the percent overlap for the hierarchical cluster analysis. We have found that this procedure results in a job taxonomy that is more consistent with our classification structure than the taxonomy resulting from a standard CODAP analysis.

Supplements to CODAP. While CODAP has significantly aided our job analysis and job evaluation research at Control Data, it is not the only software package that we use. One limitation of CODAP is the lack of flexibility with respect to cluster analysis. As a result, we make use of a software package called CLUSTER. CLUSTER is based on a series of FORTRAN clustering routines written by Michael R. Anderberg. These were written while Anderberg was at the Air Force Human Resources Laboratory (AFHRL), and his book Cluster Analysis for Applications (1973) was his Ph.D. dissertation at the University of Texas.

The Control Data adaptation of CLUSTER is extremely flexible. Options include 15 similarity indices for binary data and three similarity indices for scaled data, including the correlation coefficient, Minkowski distance, and the CODAP overlap metric. The user may also select from seven different hierarchical clustering procedures. Perhaps most importantly, CLUSTER permits

non-hierarchical cluster analysis. Thus, we can use a hierarchical cluster analysis such as CODAP's to identify the apparent number of job clusters within a sample of job holders. The results of the hierarchical analysis can then provide average job descriptions of job types as seed points for the iterative task of ensuring that each individual is grouped with that cluster to which he or she is most similar. This allows us to scrub up the clusters resulting from a hierarchical analysis. Our preliminary research leads us to believe that up to five percent of the individuals in a hierarchical analysis are misclassified--they have less similarity to the group that they have been classified in than some other cluster. Additionally, the non-hierarchical clustering method allows us to cluster far greater numbers of individuals than the hierarchical approach. Whereas our version of CODAP may cluster 2,000 cases, our non-hierarchical method in CLUSTER can group 60,000 cases. A final advantage to using CLUSTER is that it requires far less central processing time than the clustering routine of CODAP. We can cluster samples at a fraction of the cost of using CODAP.

Currently, we at CDBAI are in the process of updating CLUSTER by incorporating some additional options for improving group homogeneity. Among these is the AFFIRM algorithm developed by Schoenfeld (1970). AFFIRM identifies and excludes from analysis those job incumbents who do not fit neatly into any of the existing job clusters, or who fall very near the boundary between two or more clusters. Elimination of these incumbents tends to create more homogeneous clusters. AFFIRM then uses a non-hierarchical clustering algorithm to identify new clusters, tests to see if the incumbents who were previously excluded will fit into any of the new clusters, eliminates new outliers from the data analysis, forms new clusters again, etc., until an optimal solution is obtained. We believe that enhancements like these will make CLUSTER an even more valuable part of our job analysis system.

Another supplement to CODAP is our large-scale factor analysis program, FACTOR. FACTOR has been used to test the rational assignment of tasks to duty areas. Last year, for example, we performed a factor analysis of our software data base. The statistically derived factors closely replicated the rationally derived duty areas, suggesting the accuracy of rational duty area judgments made by job experts. We hope to use FACTOR in the future to test further the accuracy of rationally derived duty areas, and also to investigate the factor structure of knowledge, skills, and abilities (KSA's) and training competencies.

Job analysis at Control Data. To date, Control Data has surveyed approximately 9,000 employees from 16 different countries with our job analysis questionnaires. These represent management employees surveyed with our Position Description Questionnaire and non-management employees surveyed with our CODAP-based system which we call FOCAS, or Flexible Occupational Analysis System. We will complete the development of 15 non-management questionnaires for all corporate jobs by the end of 1983. Internally, we have scarcely begun to tap the potential of CODAP as a tool in the development of an integrated personnel system spanning human resource planning, selection, performance appraisal, training,

and career development as well as job evaluation.

In addition, at CDBAI, we have been assisting external clients who are interested in job analysis. To date, we have worked with ten external companies, and our current and completed external projects have so far involved almost 10,000 U.S. and international employees. With several of these firms, we have been extensively involved in the research effort, including assistance with questionnaire development, data analysis, and the ongoing implementation of the new job analysis sytem. With other clients, we have acted as a service bureau, simply assisting in the analysis of data collected by an internal research staff.

The future of CODAP. Following such a grand tribute to CODAP, it may come as a surprise to learn that we are beginning to phase out CODAP at Control Data. There are a couple of reasons for this. First, in 1983, Control Data will cease support of FORTRAN IV, which is the language that was used in programming CODAP. This will make it impossible for us to add enhancements to CODAP in the future. To convert CODAP to FORTRAN V would be a substantial effort that we will probably not undertake. Second, as part of our ongoing software development effort, we are beginning to create a number of programs that perform many of the same functions performed by CODAP. Invariably, these routines are faster and more efficient, and they do not require the huge amounts of core memory and disk storage space needed to perform CODAP data analyses. These computational programs will form the basis of a new Job Analysis Software System, JASS, which we are committed to developing at CDBAI. JASS capabilities will include large-scale cluster analysis and factor analysis via CLUSTER and FACTOR, respectively, CODAP-style analysis of task data via our new computational programs, and graphical display of major results via our new report generator programs.

References

Anderberg, M. R. Cluster Analysis for Applications. Academic Press: New York, 1973.

Gomez, L. R., Page, R. C., and Tornow, W. W. Development and implementation of a computerized job evaluation system. Personnel Administrator, 1979, 24, 46-52.

Gomez, L. R., Page, R. C., and Tornow, W. W. A comparison of the practical utility of traditional, statistical, and hybrid job evaluation approaches. Academy of Management Journal, in press.

Schoenfeldt, L. F. Affirmation Program. Unpublished computer program, University of Georgia, 1970.

Tornow, W. W., and Pinto, P. R. The development of a managerial taxonomy: A system for describing, classifying, and evaluating executive positions. Journal of Applied Psychology, 1976, 61, 410-418.

The Use of Task Based Job Analysis Data for Developing Performance Evaluation Systems

David M. Van De Voort   Beverly Stalder

Organizational Research & Development, Inc.
2455 North Star Road, Columbus, Ohio 43221

The Integrated Personnel System Concept

In 1979 the Nationwide Insurance Companies and Organizational Research & Development, Inc. commenced developing an Integrated Personnel System (IPS) in which human resource decisions and programs are data based and job related. The foundation of the IPS is a comprehensive job analysis data base. Task level data was gathered using three task inventories, each targeted to a different segment of the organization:

- an inventory for 4500 employees in 350 jobs in the range from first level manager down to, but not including, clerical jobs comprising four large job families (Administration, Claims, Systems & Data Processing, and Underwriting).

- an inventory for 1500 managers and executives designed to assess the strictly managerial content of jobs.

- an inventory for approximately 1500 employees in 500 jobs in the range from first level manager down to, but not including, clerical jobs in the 14 job families not covered by the first inventory. These 14 families are diverse in content (Legal, Facilities, Research, Planning, Personnel, Marketing, Finance, etc.) but small in number of employees (12 to 350). This inventory consists of modular surveys, each with a "core" of 242 common tasks, plus 50-200 function-specific tasks, not shared across job families.

The task inventory method was chosen for the job analysis because of the need to amass detailed information on hundreds of jobs from thousands of incumbents. Task data are amenable to computer storage and analysis and are applicable to a wide variety of personnel decisions. Incumbent ratings of time spent on tasks provide a very close link to "job behaviors" and "worker requirements", concepts that are central to legal guidelines for personnel decisions.

To date, the task database has been applied to several personnel programs at Nationwide. Evaluation dimensions for a Managerial Assessment Center were identified by analyzing the task content of "target jobs", those representing the levels and functions for which management potential is being assessed. Generic dimensions of managerial job performance derived from the managerial inventory serve as the basis for an annual supervisory rating of promotion potential. The task data base was used to develop a job classification and titling system based on similarity of task content. Requirements of entry-level professional jobs were derived as the basis for a structured employment interview guide. Task based job evaluation similar to that described by Page and McHenry, in another paper in this symposium is an ongoing effort.

FORM DEVELOPMENT PROCESS

EMPLOYEE DEVELOPMENT REVIEW

JOB ANALYSIS	:Interview representative sample of all job incumbents to gather task information about work performed.
	:Edit tasks; Assemble tasks into inventory format.
	:Administer task inventory to entire population of incumbents. Ratings of task on relative time spent and relative importance.
	:Sort or statistically derive task catagories. (Job Performance Dimensions).
IDENTIFY SUBFAMILIES	:Cluster jobs by task similarity and/or obtain job similarity judgments from job experts. Determine which jobs are to be grouped for the development of EDR inst-uments.
PRODUCE EDRs FOR EACH SUBFAMILY	:Average task ratings over individual incumbents in each job title identified as a subfamily member. Produce a list of tasks for each subfamily listed by Performance Dimension in order of relative time spent.
	:Assemble EDR from subfamily task list. Attach rating scales.

PERFORMANCE EVALUATION

JOB ANALYSIS	:As above
IDENTIFY TECHNICAL STANDARDS	:Policy makers and personnel specialists in each job family develop specific criteria for evaluating function-specific, professional/technical performance as content of a "Technical Aspects" Dimension to upplement the Performance Dimensions, identified above, which are generic to all jobs.
DETERMINE FACTOR WEIGHT RANGES	:Job family policy makers determine the range within which an individual supervisor can determine the multiplier weights for Performance Dimension ratings.

FIGURE 1

Sample Page: Employee Development Review

KEEPING WORK RECORDS

Maintain and Keep Records

DOES NOT APPLY	SECONDARY ASPECT	PRIMARY ASPECT		PERFORMANCE ON THIS TASK: NEEDS IMPROVEMENT	MEETS REQUIREMENTS	EXCEEDS REQUIREMENTS
___	___	___	Transcribe information from company forms into records, files.	___	___	___
___	___	___	Reviews and summarizes information in files, logs, records.	___	___	___
___	___	___	Maintain and update reference manual.	___	___	___
___	___	___	Organize and maintain a personal file or record system.	___	___	___
___	___	___	Maintain log of routine work activities.	___	___	___
			Additional Tasks			
___	___	___		___	___	___
___	___	___		___	___	___
___	___	___		___	___	___
			File/Organize Work Materials			
___	___	___	File applications, incomplete applications, rejected applications, claims authorization, or new files in appropriate place.	___	___	___
			Additional Tasks			
___	___	___		___	___	___
___	___	___		___	___	___
___	___	___		___	___	___

FIGURE 2

FIGURE 3

Sample Pages: Performance Appraisal

FACTOR WEIGHT

MAKING DECISIONS AND SOLVING PROBLEMS

Supporting examples:

- **Analysis**
  - Recognize problem situations and identify underlying issue(s)
  - Collect relevant information to analyze problems effectively
- **Decision Making**
  - Generate alternative solutions to problems
  - Evaluate alternatives by weighing pros and cons of each potential solution
  - Select most appropriate solutions to problems
- **Implementation**
  - Facilitate effective implementation of selected solutions to problems
  - Monitor effectiveness of selected solution to problem and make changes as needed

Extent to which individual meets job requirements (mark one)

NOT PART OF JOB	CANNOT EVALUATE	WELL BELOW		SOMEWHAT BELOW		MEETS	SOMEWHAT ABOVE		WELL ABOVE	
NPJ	CE	1	2	3	4	5	6	7	8	9

FACTOR WEIGHT

WORKING WITH OTHERS

Supporting examples:

- **Work with Others to Solve Common Work Problems**
  - Solicit opinions of others
  - Facilitate compromise by maintaining flexibility
- **Interpersonal Interactions**
  - Interact in a warm, friendly, and confident manner
  - Respond well to criticism
  - Adapt demeanor to fit the work setting
  - Provide feedback to others objectively and without personal bias

Extent to which individual meets job requirements (mark one)

NOT PART OF JOB	CANNOT EVALUATE	WELL BELOW		SOMEWHAT BELOW		MEETS	SOMEWHAT ABOVE		WELL ABOVE	
NPJ	CE	1	2	3	4	5	6	7	8	9

- 3 -

PERFORMANCE SUMMARY

1. Assign a factor weight to each dimension which represents the importance of each dimension to total job performance. THE SUM OF ALL FACTOR WEIGHTS MUST EQUAL 100
2. Once you have rated the employee's performance, transfer the performance ratings to the performance rating column below. Make sure the rating is in the row corresponding to the dimension on which the rating was made
3. Multiply each dimension factor weight by the corresponding performance rating to arrive at the Factor Performance Value
4. Total the Factor Performance Value Column. THIS VALUE SHOULD NOT EXCEED 900
5. This total Performance Value represents the employee's overall performance level and corresponds to the Overall Performance Rating Scale below

Dimension	Factor Weight	×	Performance Rating	=	Factor Performance Value
Keeping Work Records					
Planning and Scheduling Own Work					
Communicating					
Working With Others					
Making Decisions and Solving Problems					
Other Performance Dimensions					
Technical Aspects Overall Rating					
Productivity					
Total	= 100				= Total Performance Value

OVERALL PERFORMANCE RATING SCALE

Check the box which contains the individual's Total Performance Value as calculated in the table above

Extent to which individual meets all job requirements.

WELL BELOW		SOMEWHAT BELOW		MEETS	SOMEWHAT ABOVE		WELL ABOVE	
100 149	150 249	250-349	350-449	450-549	550-649	650 749	750 849	850-900

Employee Name: ____________________

Months on Current Job ____________________

Employee Number: ____________________

Period Covered by Evaluation ________ - ________

Job Title: ____________________

Date of Previous Evaluation ____________________

Job Code: ____________________

Employee* ____________________
SIGNATURE AND DATE

Evaluated by ____________________
SIGNATURE AND DATE

*Your signature here indicates that your supervisor has reviewed this evaluation with you. It does not indicate that you agree with all of the contents

Reviewed by ____________________
SIGNATURE AND DATE

- 4 -

Task Based Performance Appraisal

A key component of the IPS is the task based performance appraisal program. As advocated by Meyer, Kay, and French (1965) the two major purposes of performance appraisal are "split", with a performance review for employee development temporally separated from the annual evaluation for merit pay/promotion. Figure 1 outlines the development of the Employee Development Review (EDR) and the Performance Evaluation Form (PE). One page of an EDR is illustrated in Figure 2. Figure 3 illustrates two performance evaluation dimensions (Working with Others and Making Decisions) and illustrates how the six performance dimension ratings are multipled by dimension importance weights to derive an overall importance "score". The key aspects and contrasts of this system are summarized below:

- The PE is the traditional annual performance review with direct linkage to merit pay and promotion decisions. The EDR is solely a tool for employee development.

- The PE consists of supervisory feedback to the employee regarding performance. The EDR is a joint discussion of employee development needs and preferences. Both Supervisor and Subordinate complete the EDR and discuss their perceptions in detail.

- The EDR focus is on enhancing future performance. The PE is an evaluation of current performance.

- The EDR is a private process involving only the employee, the direct supervisor, and the supervisor's supervisor (who reviews the EDR to evaluate the supervisor's performance in developing employees). No formal record of EDRs is kept. The PE ratings become part of a central, computerized, longitudinal performance data base, which serves as a source of information for a variety of human resource planning and management decisions.

- In most cases the PE is an annual review. The EDR is a working/planning tool which employees and supervisors are trained and encouraged to use for frequent performance coaching.

- The EDR is job-related at the task level. Performance expectations are discussed in terms of specific tasks. The PE is job-related at the dimension level, consisting of ratings on a set of generic performance dimensions which are common to all jobs across the company. All employees are rated on six dimensions (see Figure 3); supervisory employees on two additional dimensions dealing with supervising, developing, and evaluating subordinates and budgeting.

- The EDR is individualized with different forms for different job families reflecting functionally specific job content. The PE is standardized, with a common set of generic performance dimensions used to evaluate all employees. A Technical

Aspect insert to the PE addresses the function specific, professional/technical competence of the employee.

To summarize the strengths of this system:

- Performance ratings are not based on personal traits but on performance dimensions which are quantitatively related to job content.

- Rating on performance dimensions and metrics which are common to all jobs permits comparison across individuals, work groups, or functions.

- Maintaining centralized performance data in computer files permits comparisons across time, thus providing built-in criteria for longitudinal "time series" tests of any organizational intervention which purports to "increase productivity" or enhance employee performance.

- Training programs can be developed to address specific task clusters where performance deficiencies are noted and training targeted to the individuals or groups with the specific need.

Finally, the implementation of this system has produced several additional benefits.

- The EDR discussion involves precise and mutual definition of job expectation on a regular basis.

- Establishment of importance weight ranges for performance dimensions by policy makers in each job family resulted in public clarification of Department-wide performance priorities.

- Performance evaluation is put on an objective basis, providing supervisors with a specific "language of tasks" to use in describing performance.

- Supervisors are, themselves, evaluated in terms of their effectiveness in using the employee development and appraisal tools, thus providing a new organization wide focus on development.

Reference

Meyer, H.H., Kay, E., & French, Jr., J.R.P. Split roles in performance appraisal, Harvard Business Review, 1965, 43, 123-129.

DISCUSSION OF JOB ANALYSIS IN THE PRIVATE SECTOR

Lt Col Jimmy L. Mitchell, Chief, OMYO
USAFOMC, Randolph AFB, TX 78150

INTRODUCTION - Since this session is running overtime, this discussion will be very brief. Overall, I am delighted with this kind of symposium. My compliments to the companies involved for sharing their job analysis experiences and data with us today. The quality of the reports was outstanding; I am very impressed with the the data displayed and by the way the reports were presented. On the other hand, as a discussant, I feel obliged to point out a few things about each of the presentations.

GAIL M. DRAUDEN, HONEYWELL, INC. - The use of work sample tests in this reseach is excellent. Likewise, the study of the transferability of people to new jobs is exciting; we have needed research into this area very much. However, we may have a language problem. Gail uses the term "task factor" to mean something quite different that what is meant by these words in most occupational reseach. Unless we have a common taxonomy of terms, we cannot be sure that we are really communicating, and thus the value of the research may be lost. This problem has been detailed elsewhere (see my Taxonomy paper, 1977 MTA Proceedings) and will not be repeated here. None-the-less, our language is something we need to be careful about. With reference to Drauden's research, one of the unresolved issues is the relationship between tasks performed and Knowledges, Skills, and Abilities (KSAs). We do not yet have an unambigious way to translate from tasks to the underlying KSAs. This is an area which needs considerably research in the future.

DAVID M. VAN DE VOORT, ORGANIZATIONAL RESEARCH AND DEVELOPMENT, INC. - This report is an excellent example of what can be done with the integrated use of task-based job analysis data for all the jobs in an organization. However, with 300 job titles and only 500 tasks, the tasks obviously must be fairly general. I like the approach used here of a "family" of inventories to cover the 14 job categories with a core of common tasks. I also very much liked the clear separation of performance review from pay and promotion actions. Someone finally noticed the Kay, Meyer, and French article and takes it seriously; this is a big plus as far as I am concerned. There may, however, be some problem in using the same tasks for all objectives. Don't be trapped into using one instrument or one level of specificity for all purposes. We may need multiple levels (possibly hierarchical levels of inventories) in order to have a flexible system for multiple purposes. I welcome the promise for a future report of more data from this research project.

STEPHEN M. COLARELLI, BALL FOUNDATION - The organizational simulation in Steve's report was an interesting and exciting new application of job analysis data. The idea of assigning persons to positions in the simulation based on their job data provides a sort of diagnostic evaluation of their present level of occupational development. You could take this idea a step further and explore better use of person and position data in a person-job match to

optimize future assignment of people to a progressive series of developmental jobs in the organization. Or, you could also use this kind of exercise to teach managers about making the best possible use of their human resources. I would commend to you Joe Ward's research on optimizing for both the organization and the individual. Some combination of what you have done with job diagnosis and his optimization scheme could have fantastic potential for dynamic career progression of human resources in private industry.

SIDNEY GAEL, AMERICAN TELEPHONE AND TELEGRAPH COMPANY - I was extremely well impressed with Sid's use of quantitative criteria (Importance ratings + 75% performing) to select tasks for evaluation. While I have some qualms about some scales (such as the Importance scale), I am familiar with his WPSS. Since it is quite like the Air Force task inventory approach, I am naturally biased in its favor. I would also applaud the AT&T research into the psychometric properties of the instruments. This is something which everybody should be doing and is long overdue. We have to know this kind of information if we are to take job analysis research seriously! I would also like to congratulate Sid on the display of his results and data. His report was extremely well done, and I wish that we had time to hear it presented in much more detail.

RON PAGE, CONTROL DATA BUSINESS ADVISORS, INC. - Ron did not leave me much time to be critical. His research is an excellent example of a historical phrase - good old "dustbowl empiricism." If it works, use it! But we also need to publish our results. Ron and his associates have tried transformations of 1.5, 1.75, and 2.0 to make time-spent data more realistic, and apparently it worked. By publishing those results, they can save the rest of us considerable research. And I think that by reporting such data here, CDBA has taken an important first step toward making their results available to the wider job analysis community. As regards the CDC development of CODAP, I applaud their enthusiasm. At the same time, I would urge some caution, in the sense of not getting so far away from the mainstream of CODAP that the evolving technologies are not useable to them. For example, task factors (difficulty of tasks or training emphasis ratings by senior technicians) have not yet been exploited in the civilian applications of the Task Inventory approach, and yet these are the areas where we in the military are getting our occupational data used more and more often in objective, real-world decision making. Overall, Ron, I enjoyed your presentation and look forward to being able to study your data in more detail.

CONCLUDING REMARKS - This has been an exceptional session; one of the best I have ever attended. There just is not enough time in a two hour session to do justice to all of the information which was made available today. I congratualate Ron Page for putting it together. We have all learned from this interaction, and I want to thank the participants for reporting their occupational research in this kind of forum. I sincerely hope that this is the start of a series of such symposia, and I look forward to seeing more such sessions at future MTA conferences, and in other meetings as well. I know that the audience joins me in saying thank you for this symposium, and a hearty "Well Done!".

ENLISTMENT AND REENLISTMENT MOTIVATION

Chair: Timothy Elig

This symposium presented current research on motivation to enter or reenlist in the military. Among topics discussed were the 1978 Selected Reserve Reenlistment Bonus Test, 1982 survey of persons entering the Army, and a literature review of motivation factors leading to military enlistment.

Enlistment Motivation in the All-Volunteer Force Environment: A Review of Major Surveys

David P. Boesel
John A. Richards
Defense Manpower Data Center

It is unlikely that any event since World War II has stimulated more military-related social research in this country than the termination of the draft in 1973. In the years immediately preceding and in the nine years since the inception of the All-Volunteer Force (AVF), military-eligible youth have come under increasing scrutiny, especially in surveys. Most notable among the surveys of this population have been the Gilbert Youth Attitude Surveys, conducted semiannually by Gilbert Youth Research, Inc. from 1971 through 1974, and the Youth Attitude Tracking Surveys, conducted semiannually from 1975 through 1980 and annually since then by Market Facts Inc. Both surveys were conducted among roughly the same population--non-prior Service, male youth in age groups 16-21 for the Gilbert surveys and 17-21 for those conducted by Market Facts. The Gilbert surveys used personal interviews, while Market Facts has been using telephone interviews. Both were cross-sectional, rather than longitudinal.

Also begun in 1971 were the DoD Surveys of Personnel Entering Military Service--that is, surveys of military recruits conducted immediately following their being sworn in at AFEES in-processing centers. The AFEES surveys were conducted annually through 1976, and once since then, in 1979. Unfortunately, the quality of the AFEES surveys has been, for a variety of reasons, inconsistent. The 1979 administration was the only carefully-controlled and fully-documented survey in the series.

Several important studies of this population--the "Youth in Transition" studies by the Institute for Social Research, for example--predate the AVF or were conducted primarily for other purposes and are, therefore, beyond the scope of our talk. Also, there have been a number of recent studies that provide excellent data on military recruits vis-a-vis their non-military peers--the Ohio State National Longitudinal Study of Youth Labor Force Behavior and the 1981 Rand Survey of Applicants for Military Service are the two best examples. Findings from both of these will be discussed in a few minutes by my colleague, David Boesel.

This review concentrates on self-reported reasons for enlisting in the military. These resemble attitudinal data and must be regarded as only one kind of variable contributing to enlistment. Others include aggregate variables such as unemployment rates and military pay, and individual variables such as parental occupation and respondent education, among others. A variety of multivariate analyses are currently being conducted in an effort to sort out the relative contributions of each, as well as the interactions among them.

The Draft-Motivated Enlistee

There was a time when one of the most common reasons for enlisting in the military was to beat the local draft board to the punch, thereby preserving an element of self-determination. In a 1964 DoD Survey of Active Duty Personnel, 20 percent of the non-high school graduates, 40 percent of the high school graduates and 58 percent of the college graduates claimed their enlistments were motivated by the draft (Lee and Parker, 1977). Yet, even in the days of conscription, another incentive showed up in study after study as far back as 1949: many enlistees and potential enlistees were strongly influenced by the opportunity provided by the Services to learn a marketable trade or skill. This motivation has persisted, and later surveys suggest it occurs within the context of a more generalized desire for self-improvement.

The Youth Attitude Tracking Survey

The Youth Attitude Tracking Survey, now conducted annually, has become a staple among the military recruiting community. This survey, the successor to the Gilbert Youth Attitude Surveys, is administered to approximately 5,000 military-eligible males (a female sample was added beginning with the 1981 survey) every fall. It provides data on enlistment propensity, attitudes toward and perceptions of the military, and a number of demographic variables. To be of maximum utility to recruiters, the sample is stratified by geographical tracking areas.

The salient findings from the last administration of the Youth Attitude Survey are summarized in Table 1.

In this table, the top seven ranked job attributes are grouped according to their achievability in military vs civilian jobs, as perceived by positive and negative propensity respondents. Positive propensity respondents are those who said they would probably or definitely serve in the military within the next few years and negative propensity respondents are those who said they would probably or definitely <u>not</u> serve. Those who expressed an

TABLE 1

PERCEIVED ACHIEVABILITY OF IMPORTANT JOB ATTRIBUTES IN MILITARY vs CIVILIAN JOBS

	MORE ACHIEVABLE IN MILITARY	MORE ACHIEVABLE IN CIVILIAN JOBS
POSITIVE PROPENSITY RESPONDENTS	• JOB SECURITY • TEACHES VALUABLE TRADE OR SKILL • DEVELOPING YOUR POTENTIAL • OPPORTUNITY FOR ADVANCEMENT	• ENJOY YOUR JOB • GOOD INCOME • EMPLOYER TREATS YOU WELL
NEGATIVE PROPENSITY RESPONDENTS	• JOB SECURITY • TEACHES VALUABLE TRADE OR SKILL	• ENJOY YOUR JOB • GOOD INCOME • DEVELOPING YOUR POTENTIAL • EMPLOYER TREATS YOU WELL • OPPORTUNITY FOR ADVANCEMENT

SOURCE: FALL 1981 YOUTH ATTITUDE TRACKING SURVEY

596-3

TABLE 2

MOST IMPORTANT REASON FOR ENLISTING IN THE MILITARY BY LEVEL OF EDUCATION

FORM 1

MOST IMPORTANT REASON	2 YEARS HIGH SCHOOL	HIGH SCHOOL GRADUATE	2 YEARS COLLEGE	4 YEARS COLLEGE	TOTAL SAMPLE
	(N=829)	(N=4242)	(N=233)	(N=149)	(N=7419)
SKILL TRAINING	25.1	29.4	17.2	21.5	26.8
MONEY FOR COLLEGE	3.0	7.0	12.4	4.7	6.8
TO BETTER MYSELF IN LIFE	42.8	38.8	39.5	43.6	39.0
SERVE MY COUNTRY	9.2	7.5	7.3	12.1	8.1

FORM 2

MOST IMPORTANT REASON	2 YEARS HIGH SCHOOL	HIGH SCHOOL GRADUATE	2 YEARS COLLEGE	4 YEARS COLLEGE	TOTAL SAMPLE
	(N=964)	(N=3980)	(N=1[illegible]8)	(N=111)	(N=7332)
SKILL TRAINING	36.3	37.7	28.8	24.3	35.4
MONEY FOR COLLEGE	7.0	9.2	17.7	7.2	9.0
TO BETTER MYSELF IN LIFE	28.3	27.3	37.9	34.2	28.0
SERVE MY COUNTRY	13.6	9.4	7.6	16.2	10.0

SOURCE 1979 DOD SURVEY OF PERSONNEL ENTERING MILITARY SERVICE

596 3

interest in military service view it as a job where they can learn, develop and advance while enjoying a high degree of job security. To achieve these job goals, they are willing to make some sacrifices in the areas of job enjoyment, income, and employer good will, which they feel would more likely be found in civilian jobs.

The negative propensity group apparently agrees that the Services are more likely to provide job security and skill training. However, they feel that opportunities for advancement and development of potential are more likely to be found in civilian jobs.

In the survey report, the authors point out that, "over time these attitudes and perceptions have remained fairly constant, though in the past year negative propensity males have come to regard 'teaches a valuable trade/skill' as one of the more desired job characteristics. This might reflect perceptions of an increasingly competitive job market and the consequent greater need to obtain practical vocational training" (Market Facts, Inc., 1982).

Surveys of Personnel Entering Military Service

With the DoD Surveys of Personnel Entering Military Service the focus shifts from eligible youths to those who have just been sworn into the Service at an AFEES in-processing center. My discussion of the AFEES surveys will be based on data from the 1979 survey.

The 1979 AFEES Survey was administered in two waves, one in the spring and one in the fall. Two forms were used with each wave, and they were modified between waves, so there were four forms in all. All eligible non-prior service males and females were sampled for a 20-day period. The two versions of the form were distributed on an every-other-person basis. There were between seven and eight thousand respondents for each form.

Respondents were asked to indicate the one MOST important reason for their enlisting in the Service. In Forms 1 and 3 the response options were listed in the same order, but for Forms 2 and 4 the order was reversed to test for order effects. The results for each pair of forms in which responses were listed in the same order were quite similar; however, a pronounced order effect shows up when results are compared across variations. This effect is particularly noticeable in the two most frequently mentioned reasons for enlisting. It can be seen quite clearly in the data for the total sample, which I've presented in Table 2 for both Forms 1 and 2. (Because of space limitations, I've omitted data for Forms 3 and 4 since they were similar to the data for their like-ordered counterparts.) In addition to data for the combined sample, Table 2 shows the top four reasons for enlistment by four educational level sub-groups.

For Form 1, SKILL TRAINING is the second most frequently cited reason for enlisting, coming after TO BETTER MYSELF IN LIFE, which appeared above SKILL TRAINING in the list of options. The results for Form 2 are just the reverse, as was the order of the response options, with SKILL TRAINING clearly leading the list. This effect is less pronounced for other reasons for enlisting.

In spite of the order effect, SKILL TRAINING makes a strong showing in these data. If you cancel out the order effect by combining Form 1 and Form 2 data, about a third of the respondents across forms said it was the main inducement to enlistment. TO BETTER MYSELF IN LIFE was also selected by about one-third of the respondents to Forms 1 and 2 combined. "TO BETTER MYSELF IN LIFE" is a rather vague statement though, and it obviously means different things to different people. It would not be unreasonable to assume that training and education comprise a major component of this concept for some people.

The data on reasons for enlistment become more interesting when their relation to level of education is examined. The appeal of SKILL TRAINING generally declines as a function of increased educational attainment. MONEY FOR COLLEGE follows a more distinct pattern, increasing in importance as a reason for enlisting for those with up to two years of college, then falling back sharply for those with four-year degrees.

Patriotism, or SERVE MY COUNTRY, is a surprisingly powerful incentive according to these data. It is the third most commonly cited reason for enlisting, and, somewhat counter-intuitively, it takes a big leap in importance for four-year college graduates.

The conclusions of the author of an unpublished paper on the 1974 AFEES Survey summarize well the major findings from this entire series. He reported that, "without any doubt, the main reason given for entering the Service was to obtain job training. This is true for all ages, races, sexes, branches of the Service and regions of the country" (Giesecke, 1976).

The National Longitudinal Survey

The Youth Cohort of the National Longitudinal Survey of Labor Force Experience began in 1979 with a sample of 12,000 youth age 14 to 22, including 1,200 military members, and has been repeated each year since then. Sponsored by the Department of Labor, with substantial contributions from the Department of Defense, the NLS Youth Cohort provides an invaluable source of information on enlistment and on the military and post-military careers of service members.

An analysis of 1980 NLS data (Kim, 1982b) emphasizes the importance of training, education, and more broadly, personal development among the reasons for enlisting given by respondents in the military. Training, a desire to better oneself, and money for college education are most frequently cited as the main reason for enlisting.

1980 National Longitudinal Survey (NLS)
Main Reason for Enlisting
(1979 Enlistees)

	% of Enlistees
Training for civilian job	28
Better myself in life	20
Money for college education	15
Travel	9
Was unemployed	8
Serve my country	7
Get away from home	6
Prove myself	5
Earn more money than on civilian job	1
Get away from personal problem	.6
Family tradition to serve	.2
Retirement/fringe benefits	.2

These data are supported by the 1981 Survey of Military Applicants, conducted as a complement to the DoD Educational Benefits Experiment of 1980. In a preliminary analysis of the Applicants data, Rand researchers find a clear relation between the probability of enlistment and the need for money to finance education: the greater the need, the greater the tendency among (in this case, high quality) applicants to actually enlist.

Enlistment Rate by Financial Need
(High Quality Applicants)

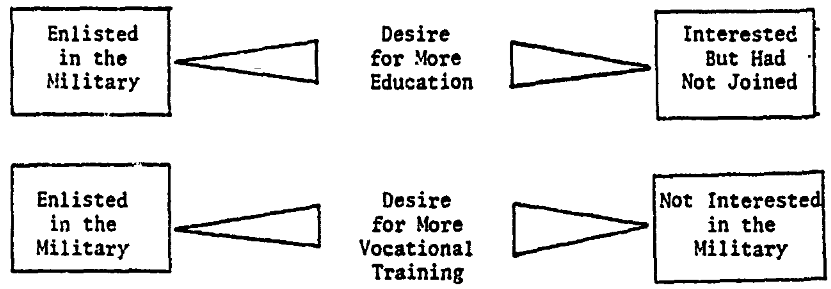
	Additional Amount Needed to Continue Education				
	$0	$1-$1000	$1001-2000	$2001-3000	$3000+
Enlistment Rate	43%	52%	59%	60%	65%
(N)	(404)	(239)	(290)	(252)	(182)

In another analysis of NLS data, Kim (1982a) assesses the role of training and education as they affect the principal choices facing high school graduates: college, other civilian pursuits, or the military. Employing multiple logit methods, Kim finds that educational aspirations and the desire for vocational training are powerful predictors of the outcome of this decision process. Male youth who have the highest educational aspirations tend to go on to college. "However," Kim notes, "when the choice is either the military or a noncollege civilian pursuit, an individual with a higher educational desire has a higher probability of choosing the armed forces," and it is a reasonable inference that a need for money to finance higher education is a factor in this decision.

The foregoing might seem to suggest that educational benefits should be increased as the supply of potential recruits dwindles in the coming years. However, Friedland and Little (1982), in an interesting analysis of National Longitudinal Survey data, argue the contrary. The authors use discriminant analysis to identify the variables which best distinguish among three groups of respondents - military members, youth who had talked to recruiters but had not enlisted at the time of the survey, and those who had never talked to recruiters.

For white males, the variable which most clearly distinguishes those in the military from those who were interested but had not joined is educational aspirations; the variable which most clearly distinguishes the military group from the not-interested group is desire for vocational training. In both cases, those in the military have the greater desire.

1979 NLS
Best Discriminating Variables:
Desire for More Education and Training



"In short," Friedland and Little observe, "variables reflecting a desire for self-improvement distinguish those in the military from those not in the military, and particularly. . .from those not in the military who considered at some time joining. . ."

Looking at the means for each group we see that the military group has the highest educational aspirations, and those who were interested but had not enlisted have the lowest. The military group also shows the greatest desire for vocational training, though the "interested-but-had-not-joined" group occupies an intermediate position between the military and the "not-interested" groups. Those in the intermediate group show a fair amount of desire for vocational training.

1979 NLS
Educational Aspirations and Desire
for Vocational Training: Mean Scores

	Not Interested in the Military	Interested but Did Not Join	Enlisted in the Military	Significance of F Test
Educational Aspiration Scale (0-5)	3.01	2.72	3.36	F .01
Desire for Vocational Training (0-1)	.58	.75	.86	F .01

From these analyses Friedland and Little conclude that increased educational benefits may be of limited value at best, because those who find them attractive have already joined, while those who expressed interest in the military but did not join are apparently not much motivated by educational aspirations. We may further note that the recent cutbacks in federal student loans and grants are likely to increase the number of youth who aspire to higher education but cannot afford college. This should improve the drawing power of the military's educational benefits program at its present level. An increased emphasis on training, we infer from the findings above, might produce some incremental gains, inasmuch as the "interested" group shows some desire for training, but again, those with the most desire are already in the Services.

While the Friedland and Little study is interesting and provocative, it leaves some important questions unanswered. For one thing, while we can see that the "interested" group has the lowest educational aspirations of the three, we cannot tell from this analysis what priority they give to education. Even if their level of aspiration is lower than that of military members, it may still be enough to warrant increased educational benefits. Second, we do not know what the youth in the "interested" group are doing and why did they decided to follow some other path. How valid is the implicit assumption that because they have talked to military recruiters they should be regarded as good potential source of accessions? Some must have applied for the military and been rejected; what proportion of the sample do they comprise? These and related questions pertaining to this significant group of youth are currently the focus of analysis at the Defense Manpower Data Center.

In summary, the specific desire for training, and a broader desire for personal development -- through training, education, and experience -- seem to be the mainsprings of motivation to enlist, according to surveys over the last decade or more. From these studies it appears that for a great many youth the Services, like the colleges and junior colleges, represent a period of maturation and preparation for adulthood.

References

Friedland, J. Eric and Little, Roger D., Socioeconomic Characteristics of the All-Volunteer Force: Evidence from the National Longitudinal Survey, 1979, Annapolis, MD, U.S. Naval Academy, 1982.

Giesecke, Lee, Results from the September 1974 Survey of Persons Entering Military Service, 1979 (Unpublished report).

Kim, (1982a): Kim, Choongsoo, The All-Volunteer Force: 1979 NLS Studies of Enlistment, Intentions to Serve, and Intentions to Reenlist, Columbus, OH, Center for Human Resource Research, 1982.

Kim (1982b): Kim Choongsoo, Youth and the Military Service: 1980 National Longitudinal Survey Studies of Enlistment, Intentions to Serve, Reenlistment and Labor Market Experience of Veterans and Attriters, Columbus, OH, Center for Human Resource Research, 1982.

Lee, G.C. and Parker, G.Y., Ending the Draft: The Story of the All-Volunteer Force, Alexandria, VA, Human Resources Reseaarch Organization, 1977.

Market Facts, Inc., Youth Attitude Tracking Study: Fall 1981, Arlington, VA, 1982.

AD P000838

THE 1982 DA SURVEY OF PERSONNEL ENTERING THE ARMY

Timothy W. Elig, Paul A. Gade, and Joyce L. Shields

Paper presented at the meeting of the
Military Testing Association
San Antonio, Texas
November 1982

U.S. Army Research Institute for the Behavioral and Social Sciences

5001 Eisenhower Avenue
Alexandria, Virginia 22333

Commercial: (202) 274-8275
Autovon: 284-8275

The authors acknowledge the cooperative efforts in fielding the 1982 DA Survey of Personnel Entering the Army of CPT Larry Davis of the Human Resources Development Directorate (ODCSPER. Department of the Army) and Mr. Richard Thompson of the Soldier Support Center--National Capital Region (TRADOC). Mr. Thompson's efforts in developing drafts of the survey are greatly appreciated. This survey effort could not have succeeded without the cooperation of the personnel of the US Army Reception Stations; their efforts in administering this survey are greatly appreciated. We also wish to thank Dr. Zahava Doering of the Defense Manpower Data Center for sharing her expertise. Dr. Doering's review of early drafts of our 1982 DA Survey were very helpful. The authors of course accept full responsibility for the final content of the survey and of this paper.

THE 1982 DA SURVEY OF PERSONNEL ENTERING THE ARMY

Timothy W. Elig, Paul A. Gade, and Joyce L. Shields

U.S. Army Research Institute for the Behavioral and Social Sciences

The purpose of this paper is to document the 1982 DA Survey of Personnel Entering the Army and to introduce this survey to the military manpower community. Because of the brevity required for this forum, it is impossible to do justice to the full scope of results obtained from this survey effort. Thus we focus in this paper on the development and administration of the survey rather than on the results. The results included in this paper are meant only to be suggestive of the possible scope of questions which can be addressed in surveys of this type.

Background

The 1982 DA Survey of Personnel Entering the Army was developed to answer questions concerning the demographics and enlistment motivations of new Army recruits. Military personnel planners require such information to monitor current recruiting strategies and to forecast future enlistment and reenlistment trends. While there is an apparent need for such information on a regular and timely basis, we know of no effort to collect such information on a regular basis. Prior to the current study, the most recent effort to collect demographic and motivational data from a large sample of military recruits was conducted in 1979 by the Rand Corporation (Doering, Grissmer, & Morse, Notes 1 & 2).

Military recruiting in 1982 is dramatically changed from military recruiting in 1979. While Army recruiting in FY79 suffered one of the poorest years in both quantity and quality since the inception of the All Volunteer Force (AVF), the high quality of FY82 Army recruits with no loss in quantity is unprecedented under the AVF. Army personnel policy planners need to know who these 1982 recruits are and why they decided to enlist. This knowledge should facilitate efforts to capitalize on the current surge in high quality applicants.

Survey Development

The 1982 DA Survey of Personnel Entering the Army is almost wholly based on the 1979 DoD Survey of Personnel Entering Military Service (Doering, Grissmer, & Morse, Notes 1 & 2). Questions were selected from the 1979 DoD Survey, and in some cases modified, to fit the purposes of the 1982 DA Survey. In taking this approach we gained two major advantages. First of all, by using previously tested items we avoided the neccessity of a long developmental effort to insure items appropriate for the subject population. The other major advantage of this approach was that it insured the availability of a cross-sectional comparison group in the Regular Army recruits surveyed in 1979.

The views expressed here are those of the authors and do not neccessarily reflect the views of the Department of the Army.

The 1982 DA Survey was designed to collect information about enlistment motivation and personal background. Motivation for enlistment was assessed both directly and indirectly. Direct questions included 11 true-false items on specific reasons for enlistment and 2 forced choice items in which respondents were asked to indicate the Most and Second Most Important reasons for enlistment. Indirect assessments of motivation for enlistment were based on personal background questions, on educational history and aspirations, financial and employment history, and on family history as well. In addition, questions about demographic topics such as gender, ethnic group, marital status, and rural-urban background were included.

A major source of supplemental information has been planned for the 1982 DA Survey. Individual survey responses have been matched by SSAN to computerized accession records. Thus we are able to look at survey responses segmented by variables such as AFQT, length of the enlistment contract, and enlistment/educational bonuses received.

Survey Procedures and Administration

The 1982 DA Survey of Personnel Entering the Army was administered to recruits in group settings during initial entry processing. For the first survey period, 3-7 May 1982, all recruits processing through five of the seven US Army Reception Stations were surveyed. Because of a conflicting mobilization exercise, the other two stations were not available until the end of May. All recruits processing through all seven stations were surveyed during the periods of 24-28 May and 14-18 June.

A total of 3313 new Regular Army recruits were surveyed with the 1982 DA Survey of Personnel Entering the Army.* This is approximately 3% of all FY82 Regular Army accessions. Ninety-five percent of these survey respondents (3155) were matched by SSAN with their accession records in the AFEES Reporting System (ARS).

Depending on local conditions, the Personal Affairs Branch or Testing Branch at each Reception Station administered the surveys in accordance with written procedures prepared by ARI and the Soldier Support Center-National Capital Region. ARI personnel were in close telephonic contact with local personnel throughout the survey period and visited each Reception Station during the first week of the survey to observe the administration conditions and procedures. Local variation in procedures appeared to be minimal. However, the possibility of sample biasing did arrise at the Ft. Jackson Reception Station during the second and third weeks of the survey. This station requested and received permission to sample recruit companies rather than survey everyone being processed at the station. This exception was granted because an unusually large number of recruits was being processed by the station at that time which required extremely tight scheduling of recruit and station personnel time. Station personnel were instructed to survey by company and to favor infantry

*US Army Reserves and National Guards account for 1660 and 2812 respectively of the total survey sample of 7785. Further information on these samples are available from the authors.

companies in the selection process. Over... ling of infantry companies was done to compensate for the absence of infantry recruits surveyed during the first week of the survey. The infantry reception stations had been unable to survey because of the conflicting mobilization exercise. The result of this sampling stategy was a slight oversampling of infantry recruits.

All Regular Army, Army Reserves, or Army National Guard recruits were requested to complete the survey. The Privacy Act Notice printed on each survey was read to all personnel. The voluntary nature of participation in answering each question was emphasized. Only a few individuals would not answer any particular question; however, many individuals would miss at least one question.

Results

Survey Representativeness

The sample size, our success in matching cases with ARS records, and the low rate of nonresponse are all positive signs that this survey effort has succeeded in capturing useful data about attitudes and motives that influence enlistment decision making. However, there are several aspects of the survey procedures that must be considered when interpreting the results. The usefulness of this data base lies much more in representing segments of the market rather than in a representation of all FY82 Army enlistments.

The survey sampling covers only the third quarter of FY82. The impact of regular seasonal variation or other shifts in motivational patterns during the course of the year are not accounted for in results reported here. The possibility of seasonal bias is attenuated somewhat by the fact that we are dealing with accession rather than contract data. People who are included in our survey signed enlistment contracts at various times of the year under the Delayed Entry Program (DEP). As can be seen in Table 1, over half of the sample contracted for enlistment at least one month prior to accessioning.

Table 1

Percentages of Respondents Signing Enlistment Contracts by Month

	FY81				FY82								
Month In Which Contract Signed	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN
Percent of Sample	7	7	4	4	4	3	5	4	4	10	29	14	5

Note: n = 2700 Non Prior Service Regular Army recruits.

The DEP has made enlistment decision making a complex process of multiple decision points. For people who enlist in the DEP, enlistment decision making involves at least a decision to sign a contract and a

decision to fulfill the contract and access. Our respondents were asked to report reasons for enlistment based on memory. The reasons given by our respondents for contracting are probably confounded with their reasons for accessing.

Thus the results of our sample of 3rd quarter FY82 accessions are best interpreted as indicative of the relative strength of motivations for enlistment in FY82 rather than as definitive of actual percentages of FY82 accessions motivated in certain ways. The major strength of this survey is in defining the motives of specific market segments. For example, this survey can be used to study the characteristics of recruits motivated by a desire to fund a college education. The timing of this survey is particularly good for the comparison of the motives of recruits recently graduated from high school with the motives of other recruits. This comparison is of particular importance for the Army Recruiting Command's efforts to penetrate the high school market.

Most Important Reasons for Enlistment

The 1982 DA Survey of Personnel Entering the Army asked each respondent to indicate the first and second most important reasons for their enlistment. The only difference in these questions from the questions asked in the 1979 DoD Survey is that one reason (Travel) which was asked in 1979 was not asked in this survey. One of the eleven reasons in the 1979 Survey had to be dropped for the question to fit the 10-answer format used in the 1982 DA Survey; travel was selected as being of least interest to our current research.

Table 2

Reasons for Enlistment of Non Prior Service Regular Army Respondents in 1982 DA Survey and 1979 DoD Survey

	PERCENT OF RESPONDENTS		
Which one of these reasons is your MOST IMPORTANT REASON (SECOND MOST IMPORTANT REASON) for enlistment?	1979	1982	
	MOST IMPORTANT	MOST IMPORTANT	SECOND MOST IMPORTANT
I was unemployed	4	10	11
To be away from home on my own	5	4	7
Chance to better myself	39	30	20
Travel (not measured in 1982)	4	--	--
To get away from a personal problem	1	1	2
To serve my country	10	9	10
Earn more money	1	2	7
Family tradition to serve	0.5	1	2
To prove that I can make it	4	6	8
To be trained in a skill	26	22	18
Money for a college education	7	15	15
	100	100	100

Table 2 lists the percentages of respondents in the current survey who chose each reason as the first and as the second most important reason for enlistment. Notice that the rank order of reasons is the same for the first and second most important reasons in the 1982 sample. Chance to better myself, skill training, money for a college education, and unemployment were selected as most important or second most important by 49%, 40%, 29%, and 20% of the 1982 sample, respectively.

The first column of Table 2 lists the responses of recruits from the spring wave of the 1979 DoD Survey. These recruits were surveyed in March and April of 1979 after signing enlistment contracts. Note that data are not included here from respondents who were given the reasons in the opposite order because order effects were found for this question in the 1979 DoD Survey (Boesel & Richards, Note 3). The rank order preference for the two top responses was reversed when their ordering in the list was reversed in alternate forms of the survey. This reversal is most likely to occur between pairs of reasons of almost equal attractiveness to recruits.

The two top rated reasons in both 1979 and 1982 are a chance to better myself and skill training. Some clues as to why 1982 was a better recruiting year than 1979 can be seen in the reasons that gained in importance in 1982 compared to 1979. Money for a college education and unemployment are the two reasons which show major changes between 1979 and 1982. Either or both of these could be associated with the general

Table 3

Reasons for Enlistment of Male High School Diploma Graduates

		PERCENT OF RESPONDENTS					
		BY AFQT				BY GRADUATION	
Which one of these reasons is your MOST IMPORTANT REASON (SECOND MOST IMPORTANT REASON) for enlistment?		I, II n=837	IIIA n=439	IIIB n=624	IV n=666	RECENT n=1441	NOT RECENT n=1110
I was unemployed		9	9	10	12	F 8	13
To be away from home on my own	B	2	4	6	5	D 5	3
Chance to better myself	C	21	33	32	33	F 25	35
Get away from a personal problem		1	0.5	2	1	1	1
To serve my country	A	9	9	6	11	E 10	7
Earn more money		2	2	3	4	3	3
Family tradition to serve		0.1	0.9	0.6	0.3	0.7	0.5
To prove that I can make it		6	5	7	6	D 7	5
To get trained in a skill	A	24	21	26	20	21	24
Money for a college education	C	27	16	8	7	F 19	10
		100	100	100	100	100	100

Note: NPS Regular Army Recruits only.
A: $p < .05$ for AFQT D: $p < .05$ for GRADUATION
B: $p < .001$ for AFQT E: $p < .001$ for GRADUATION
C: $p < .0001$ for AFQT F: $p < .0001$ for GRADUATION

improvement in recruiting quantity. However, when we examine analyses segmented by quality, these reasons do not appear to be equally responsible for the quality gains in 1982.

Male high school diploma graduates are the prime market segment for the Army Recruiting Command (USAREC). In Table 3, we focus on this market segment. Table 3 presents the most important reasons for enlistment of male NPS Regular Army recruits whose highest educational certification is the high school diploma. Respondents are segmented in this table separately by AFQT and by when they graduated from high school. Recent graduates are those who graduated in January 1982 or later, and either enlisted as high school seniors or within 3 months of graduation.

The prime Army recruits in this survey appear from the data in Table 3 to be more motivated by educational incentives than by unemployment. A male HSDG recruit is more likely to be enlisting for money for a college education the higher his AFQT category is; by these self-reports a male HSDG in Category I or II is more likely to enlist for educational benefits than for any other reason. Seniors in the classes of 1982 were more likely to enlist for educational incentives than the other male HSDG recruits in the sample who had enlisted after leaving school. High school seniors were less likely to say they enlisted because of unemployment or to better themselves but were more likely to report serving the country as the most important reason for enlistment.

Discussion

The results reported here show the need for routine collection of information on enlistment decision making and the need for careful market segmentation of the data. The Army Research Institute (ARI) has undertaken an ongoing research effort to collect and analyze both longitudinal and cross-sectional information at key points in the enlistment decision process of Army recruits. The 1982 DA Survey of Personnel Entering the Army reported here has already been revised to include questions on advertising and more questions on motives and economic incentives. Separate forms of the revised survey for active duty and reserve Army forces have already been administered.

Reference Notes

1. Doering, Z. D., Grissmer, D. W., & Morse, J. S. 1979 DoD Survey of Personnel Entering Military Service: Wave 1 user's manual and codebook (Rand Note N-1605-MRAL). Santa Monica, Ca.: The Rand Corporation, November 1980.

2. Doering, Z. D., Grissmer, D. W., & Morse, J. S. 1979 DoD Survey of Personnel Entering Military Service: Wave 2 user's manual and codebook (Rand Note N-1606-MRAL). Santa Monica, Ca.: The Rand Corporation, December 1980.

3. Boesel, D. P., & Richards, J. A. Enlistment motivation in the All-Volunteer Force environment: A review of major surveys. Paper presented at the meeting of the Military Testing Association, San Antonio, November 1982.

ISD SYSTEMS APPROACH TO TRAINING IN THE PUBLIC SECTOR

Chair: Stewart Malcolm

There is little doubt that the Instructional Systems Development (ISD) model and procedures such as the Inter-Service Procedures for ISD (ISISD) are a comprehensive basis for development of training programs. This basis has been established by many varied and intensive research efforts to ensure the validity and accuracy of the procedures. However, for the most part, the model and associated procedures have been developed for military organizations which have significant supporting agencies as well as large populations, which makes the entire effort cost-effective. This discussion focused on the ISD model and the principles which are essential to maintaining system integrity. Discussion included methods and procedures which could be used to ensure that a systems approach was achieved but with less complex and resource intensive methods and procedures. Individual experiences of panel/audience members in introducing a systems approach to training in their organization were highlighted. While recognizing the tremendous value of the ISD model, many public and private sector agencies must compromise on the approach to training development due to cost/resource constraints. The principles and procedures that seem to be compromised the most are those associated with Analysis.

AD P000839

SYSTEMS APPROACH TO TRAINING - A MANAGEMENT MODEL

Stewart P. Malcolm and Ian L. Jackson
Quality Assurance Division,
Staff Development Branch,
Public Service Commission
Ottawa, Ontario
Canada K1A 0M7

INTRODUCTION

This paper describes the progress to date in the design and implementation of a model for a systematic approach to training, such as an Instructional Systems Development (ISD) type system, across the civilian departments of the Canadian government. Our experiences will be of interest to other organizations that include a number of heterogeneous operations and functions and use training as a human resource management tool.

Currently, the management of training ranges from completely systematized operations in a number of departments to haphazard, informal applications in other departments.

This paper covers:

- the background to the project
- the analysis and development of system specifications
- the resulting system and supporting manual
- implementation
- conclusion

The project team experience included military and civilian applications of systematic and unsystematic approaches to training.

BACKGROUND

Organizational Roles (Simplified)

- The Treasury Board of the Canadian Public Service sets and approves policy, procedures, budgets, standards, etc. for the training function across the public service
- The Departments and Agencies are delegated the responsibility to operate training by the Treasury Board
- The Public Service Commission provides training services to the Treasury Board (e.g. accreditation of instructors) and to the Departments (e.g. over 100 different courses)
- The Auditor General is responsible for auditing the accounts of the government departments and agencies

Needs Identification

The need for a common training system that supplies a rational framework for the management of training activities and resources has been identified through a number of sources:

- two independent studies identified growing irrelevancies in training and weaknesses in the management of training resources

- the Auditor General identified weaknesses in training policies and controls

- senior government officers have pointed out missing elements such as performance standards for instructors/trainers

- increasingly scarce resources have resulted in a closer scrutiny of training cost-benefit ratios by policy and planning decision makers

Results

To meet the identified needs, the Treasury Board issued a new training policy that:

- required training to become work-performance oriented

- imposed stronger control mechanisms (e.g. comprehensive audit, planning, reporting, etc.)

- started the development of mandatory accreditation program for federal trainers/instructors

- established a Staff Training Council of senior officers from the Treasury Board, Public Service Commission, and a cross section of Departments to "manage" the training function

The Staff Training Council, in turn, directed the Public Service Commission to develop a Systems Approach to Training model and manual that would become the basis for rationalizing training activities and decisions.

ANALYSIS AND DEVELOPMENT OF SYSTEM SPECIFICATIONS

Analysis

The Analysis reviewed:

- current systems operating in Public Service departments and agencies

- the relationships between training and non-training managers

- current systems operating in other organizations e.g. military

- recent literature on the subject of systems in training and their application

- the opinions and ideas of experienced trainers and non-training managers

Specifications

As a result of the analysis, the system specifications had to include:

- The establishment of common straightforward terminology. A serious breakdown in communications has occurred in many areas between training technologists and non-trainers because of the proliferation of training jargon. In some cases the breakdown was occurring among training technologists working in different functions such as design and evaluation. A recent article hi-lited more than 20 different terms referring to training objectives alone. To quote the 1982 MTA keynote speaker, Major General Armstrong "Tell me in agricultural terms". The 1980 MTA keynote address by Lieutenant General Carswell called for clearer communications and an active cooperation with the decision makers.

- A management rather than training technology orientation. In addition to the communication problem, training has been allowed to drift away from the operations and become more of an isolated staff function. As a result client managers do not always manage training and in some cases are not involved in the training process. As a result, there have been resourcing problems and shortfalls in operationally controlled activities such as selection of participants and on-the-job training.

- Flexibility. The system has to be applicable in a variety of organizations as diverse as Corrections (prisons) and External Affairs (embassies). The system has to be able to cover projects ranging from a performance problem with one or two tasks to an occupational group of several thousand workers in a number of jobs. The system must recognize the availability of time and resources as the starting point.

- A comprehensive audit trail. Training is one of the few remaining functions without a specific comprehensive audit guide that supplies the framework for management and operational audits. The training system and supporting manual has to supply a description of the basic activities and decision points for the audit guide.

- Non-threatening. The system has to be seen by the departmental managers and their training technologists as a "good" thing. That is, the departments with systems in place will not have to make significant changes while other departments should find the system to be a guide and aid for system implementation.

- Environmental Emphasis. The system had to place greater emphasis on the environmental factors that are not specific to particular task(s) or jobs but overlay the entire operation. For example, interviewing in a manpower office with an unemployed person is not the same as an immigration officer interviewing a possible illegal immigrant. Training must take into consideration environmental factors such as policies, physical location, cultural differences, etc.

THE SYSTEM AND MANUAL

Our Systems Approach to Training (SAT) is made up of five phases - ANALYSIS, TRAINING DESIGN, EVALUATION DESIGN, CONDUCT, and VALIDATION, in order to facilitate description.

Key elements

- The manual describes a system that is a framework for the management of training activities and resources. It is not addressed specifically to the training technologist but rather to the user.

- The system is a prototype that will run more efficiently and effectively when customized to specific projects in particular organizations. The prototype represents minimum standards only.

- When documented, the system provides the basis for a comprehensive audit trail

- Training technologists cannot activitate the system or its phases. The system must have a sponsor from the department's management cadre with the necessary authorities.

- Wherever possible, operators of individual phases are limited in their own standards, e.g. The Analysis phase produces the training objectives for implementation by the Training and Evaluation design phases.

- The system encourages the use of subject matter experts from the sponsor's organization on a temporary basis as the key resource, wherever practical. The objective being to integrate training as closely as possible into operations.

- The manual is easy to read with a minimum of jargon. It is also unbalanced in emphasis to address specific weaknesses in current systems as well as the non-trainer's understanding of training. The target level of literacy was a grade 10 equivalent.

IMPLEMENTATION

In the Public Service Commission, we expect to convert the current catalogue of courses to a systems base in the next three to five years. The resource situation dictates the conversion schedule. Meanwhile, all new courses are

being developed to conform to the system principles. We are in the process of developing format models based on the experience of implementing the phases. Other implementation activities, in process, include:

- A multi-disciplined working group is developing a comprehensive audit guide for use in all federal departments. Completion is planned for a pilot in March, 1983.

- Departments are being briefed on adapting the system manual to make it organizationally specific.

- Performance standards and mandatory training technologists accreditation courses are being developed, based on the system parameters, for implementation in 1982.

CONCLUSION

The results of the initiatives taken to date will not be available for a year or two. But we feel that our approach in developing the system as a management process will meet Canadian Federal Government needs as long as we keep it both pragmatic and practical. We are concerned with the art of the possible. It is also obvious that the systematic approach will be in a state of development requiring continuing adjustments for a number of years. Our efforts will be directed at keeping the training system aimed at specific work being performed by defined workforce in a particular organization. A limited number of copies of the manual are available on request.

PAPER PRESENTATIONS

CHAIR PERSONS OF PAPER SESSIONS

NAME	SESSION TITLE
Antons, Chris	Training Tests
Archer, Wayne	Air Force Occupational Research
Arnott, Gail	Inovations in Testing
Ballard, John	Attrition and Counter Attrition
Black, Doris	Training Research
Buescher, Ruth M.	Performance Appraisers
Clark, Harry	Criterion-Referenced Tests
Cowan, Doug	Training Issues
Edwards, John	Vocational Interests and Aptitudes
Falle, Ian	Systems Design and Evaluation
Fracker, Lee	Tactics Training
Gentner, Frank	Tactical Symbology
Gott, Sharon	New Approaches in Personnel Measurement
Gould, Bruce	Training Evaluation
Jansen, Hans	Navy Occupational Research
Knight, Ralph	Readiness Evaluation
Lipscomb, Suzanne	Training Requirements
Phalen, Bill	CODAP 80
Ribera, Rick	Psychological Screening
Roach, Ben	Selection & Classification Testing
Ruck, Hendrick	Occupational Studies
Short, Larry	Stress Management
Stanley, Paul P.	Testing Theory and Models
Stephenson, Stan	Special Topics
Tartell, Jay	Inovations in Job Analysis
Thompson, Nancy	Performance Appraisal
Weeks, Joe	Enlistment Standards
Welsh, John	Recruiting
Wiekhorst, Linda	Officer Training Issues

AD P000840

ANALYSIS AND EVALUATION OF COMMAND GROUP OPERATIONS

Authors: Major Denis Aitken MSc and Major Robert R. Begland PhD
Training Consultants, Army School of Training Support, England

SUMMARY

This paper reports on a study carried out on behalf of the Director General of Army Training (UK) to evaluate the existing Battle Group HQ trainers in order to enhance them and to assess their utility for training command groups at higher levels.

A detailed analysis of the workings of a Battle Group HQ was carried out by the study team making use of a technique labelled "Scenario Analysis." Evidence suggests that this methodology may be of value when producing collective training objectives for similarly complex command groups.

The paper describes the analysis methodology used and briefly reviews the evaluation results.

BACKGROUND TO THE DEVELOPMENT OF BATTLE GROUP TRAINERS

The rationale behind the development of the present Battle Group Trainers (BGTs) arose from the complexity of the modern Battle Group as a fighting unit, the command of which in war would be a formidable task. The difficulty of this task would be increased by the reducing opportunities for the components of the Battle Group to train together in the field in peace because of lack of deployment space and because of financial and equipment constraints. It followed that the maximum benefit had to be obtained from limited collective field training. To obtain this benefit optimum use needed to be made of preparation for these exercises and, in addition to the usual study days, model exercises and TEWTS, exercises were also required in command and control, communications and tactical decision-making without the need for the costly deployment of troops in support.

DESCRIPTION OF THE PRESENT BATTLE GROUP TRAINERS

The BGTs were designed to subject the commanding officer and his staff to stressful situations, relating exercises to real ground and portraying battle conditions which are as realistic as possible within a realistic time frame. The training system can best be described as a combination of a tactical exercise without troops (TEWT), a command post exercise (CPX) and a war game. The trainers have three main components; the BGHQ, the control room and a tactical HQ.

The BGHQ is very much like a theatrical set. It uses actual vehicle bodies and realistic mock ups to portray the situation of a BGHQ which has taken up temporary accommodation in an old farm and outbuildings. The "set" produces, as much as is possible, the conditions that the commanding officer and his staff would expect to work under in the field, in terms of varying light levels and the background noise of battle. The information which flows into the HQ through simulated radio networks is generated in the control room.

During the TEWT phase of the training period the commanding officer makes his plan, issues his orders and co-ordinates compliance with his plan. All of this activity takes place over real ground near the trainer. The second phase of the exercise is to "fight" the battle using a computer supported simulation in the control room. The main tool used during the simulation is a master map board. The map (scale: 4cm represents 100m) is very detailed, showing features down to individual hedgerows and ditches. Own and the enemy's men, individual vehicles and major equipments are represented by symbols on the map. Contouring is shown by coloured layering and, as an aid to intervisibility, valleys and ridges are highlighted by coloured lines. Around this map sit the commanders of all the sub units involved in the battle (company commanders, reconnaissance troop commander, etc). These people "fight" the battle in respect of their own troops and report to, and seek guidance and assistance from BGHQ staffs via the simulated communications system. The play of the battle is free to the extent that the actions and reactions in the battle depend, on the one hand, on the commanding officer's plan and the way he and his subordinate commanders carry out this plan and on the other hand, upon the result of each and every individual sighting, engagement or movement.

The computer sub-system is used to generate the combat information by assessing the likelihood of sightings, of detection and the outcome of engagements between opposing units. Engagements are assessed using a data base of hit and kill values which take into account the range and circumstances of opponents as well as the weapon types. Engagements can be as simple as one on one, one on many or many on many. The computer sub system also accounts for ammunition, records battle casualties and assesses the effects of artillery, mortar, helarm, fighter ground attack and air defence weapon systems. The simulation is therefore used to generate the information used by the BGHQ to control the battle and is, except in the circumstances described for the tactical HQ, closed to the BGHQ staff.

The tactical HQ is a simulation, again using vehicle bodies and theatrical set, which allows the commanding officer to leave his main HQ during the battle. All the normal communications are provided and a combination of slides and close circuit television, focussed on the map board, is used to allow the CO to see the battlefield as he "moves."

Overall the environmental realism is achieved by theatrical means whereas the realism of the information generated during the simulation is a consequence of the attention paid to timings, movement, intervisibility, the chance of sightings, the application and assessment of direct and indirect fire, accounting for ammunition and battle casualties and the use of realistic enemy tactics. The emphasis during the design process was therefore on environmental and information realism. However, the training benefits which accrue from simulations are dependent upon task fidelity as much as realism. No matter how scientifically the information is generated during the simulation process, if the information is being used in inappropriate ways and causing the development of disfunctional behaviour then the expenditure of resources on the training system is inefficient.

BACKGROUND TO THIS STUDY

A number of suggestions for the future development of tactical training

simulators had been considered as early as February 1981. Included in these was the possible requirement for brigade and divisional level trainers, particularly in the continuing context of increased financial constraints, pressure on training areas and equipment restrictions. The use of simulation for training commanders and their staff at all levels was therefore becoming increasingly attractive and higher level trainers seemed a logical extension of the BGT concept.

While suggestions for the development of tactical training simulators were being discussed, DGAT was aware that the concept of such a training system had not yet been validated. To embark upon the development of tactical training simulators at higher levels without taking into account the lessons learned during the development of the BGTs and without firm evidence of their effectiveness could lead to inefficient development. Furthermore, the BGTs themselves would require replacement of their hardware in the mid 80s, and it was clearly sensible that the opportunity should be taken to review possible system enhancements. Consequently it was essential that a prerequisite for any developments of the BGTs themselves, or of trainers at any levels, would have to be an evaluation of the BGTs already in operation.

The terms of reference for the present study were for the Army School of Training Support (ASTS) to carry out:

a. An evaluation of the BGTs at Bovington and Sennelager, taking into account their expressed aims and acknowledged limitations.

b. An investigation into possible ways in which the BGTs could be developed to improve their effectiveness, taking into account relevant developments in commerce, industry and the Armed Forces of other nations.

THE PROBLEM

Traditionally the Commanding Officer of a battalion or regiment has been given a great deal of freedom to run his unit in his own way. The emphasis in any evaluation of the units, and consequently the CO's, readiness for combat has been summative. Hence, although guidelines are provided in field manuals and personnel are established to man the HQ in accordance with the guidelines, there are as many variations on the theme as there are commanding officers. The organisation of the BGHQ is further influenced by the operational role of the unit providing the HQ. Armoured regiments are established differently from mechanised battalions in terms of personnel and vehicles. A move from the British Army of the Rhine to a station in the United Kingdom will also cause changes in emphasis. Finally the organisation of the BGHQ reflects the CO's own experience and his perception of the level of training of the individuals who become part of his HQ on operations. The tasks to be carried out in a BGHQ, both individually and collectively, are therefore a function of the organisational structure of the HQ, the level of training of the individuals who constitute the HQ staff and the operational role of the battle group. In terms of any individual within the HQ, the tasks he will have to perform will be an amalgam of those things demanded of him by the CO through giving him in a practical role within the HQ, the tasks forced upon him by his involvement in a particular type of operation and his background and training.

The study team had specifically and appropriately been tasked with reviewing the objectives of the BGTs. However, the preliminary investigation

phase had revealed that no such objectives existed. In order to conduct a vigorous and objective evaluation of the trainers as possible an essential but major stage would be the listing of the OPERATIONAL activities of a BGHQ and its supporting staff.

THE CRITERION PRODUCTION PHASE

The reason for this phase was the identified need to produce a set of criteria, or a base line, against which the BGTs would be evaluated objectively. Initial liaison with various training establishments revealed that there was no document available which listed those activities that might be expected to be undertaken by a BGHQ while on operations in NW Europe.

In order to produce this list of operational activities an approach on three fronts was adopted:

a. A "Training" perspective; in which the skills and knowledge that could be expected of a BGHQ staff were obtained from Training Establishments.

b. A "User" perspective; in which the views of a number of present and past COs of Battle Groups were obtained concerning the operational functions of a BGHQ.

c. An "Operational" perspective; in which a major operational scenario (the Defence) was analysed together with a number of special contexts in order to identify the functions of each member of the BGHQ and supporting staff which could be expected to take place while on operations.

This paper will concentrate on the Operational Perspective. This involved the team in analysing the activities of a BGHQ at a more detailed level than had been possible in the "User's Perspective." To obtain this detail a fresh technique was adopted, called "Scenario Analysis." The BGT staff of both trainers were asked to help in the provision of this detail. The combined staff not only represented a multi arm perspective, but because of their own prior training and experience at BGT were particularly well qualified to consider the operational activities of a BGHQ.

SCENARIO ANALYSIS

This analysis involved joint activity and agreement by both these multi arm teams working chronologically through a series of BG operations with a member of the study team. As a first step in the analysis the team identified those contexts likely to cause significant activity in a BGHQ. In the Defence Scenario twenty-five such contexts were identified. The second stage of the analysis was to agree which BGHQ members had significant input to that context. The number of appointments identified varied according to the context. The third step in the process was to determine the activities of each identified appointment. An example of the activities of the Mortar Officer identified in the context of COMBAT TEAM (CT) OVERRUN is given below in the format of the analysis proforma used:

CONTEXT - Combat Team Overrun

INPUT	ACTION BY	EVENT DESCRIPTION
'A' CT reports 'B' CT's posn has been overrun	Mor	BG reacts to one of its own sub-units being overrun

Desired Response:

1. Implement DF FP	4. Request Ammo resupply
2. Receive SITREP from MFCs and pass to BC	5. Discuss changes to FP with BC if required
3. Move Mor Base Plate if threatened by breakthrough	6. Reallocate MFCs where appropriate

As Measured by:

1.	Accurate and timely response to FP.
2.	Mors report ready in new loc.
3.	MFCs ack changes to FP and new orders where appropriate.

A similar process was undertaken for each of the other eleven BGHQ members previously identified as being involved in this context.

The study team was extremely grateful to the staff of the two BGTs who contributed to this analysis, which was both time consuming and difficult. Agreement, which was important and fundamental to the end result, was not always easy to achieve due to procedural or organisational differences between the arms. At the end of this analysis both BGT teams remarked that their own knowledge and understanding had grown considerably because of the requirement to consider the workings of a BGHQ together, systematically, in detail and achieve agreement.

THE EVALUATION PHASE

The Criterion Production phase was completed immediately before Christmas 1981 and by combining and comparing the data from the various perspectives a comprehensive task list was produced of the activities a BGHQ could expect to undertake on operations in NW Europe. This task list of operational activities became the basis for the questionnaires which were the main information gathering technique for the evaluation phase. While time precluded the obtaining of formal approval by any authority, informal but informed comment from trainers and ex commanders indicated that this task list was both comprehensive and technically sound. Apart from minor comments, it subsequently received professional and technical endorsements from all the units to whom it was circulated, and a number of units, both regular and TA asked for copies as an "aide memoire" for their collective training.

Apart from MOD Command and Training Establishments, a total of 45 field units were approached and provided information for this study.

This represented a majority of the Battle Group and battalion sized units using BGTs. During the evaluation phase information was obtained from appropriate sources on the following:

a. Opportunities provided by the BGTs for carrying out the operational activities identified using "Scenario Analysis" and the opportunities taken by users.

b. The importance of the activities on a five point scale ranging from "Absolutely Critical" to "Not Important."

c. The relative training value of the BGT compared with that of FTXs and CPXs for each of the identified activities.

d. The overall context of BGT training, in order to ascertain where attendance at BGT should occur in the training cycle.

e. The fidelity of the various tasks as presented at BGT.

f. The reality of the environment at BGT in which the tasks are presented (This was termed "Physical Fidelity").

g. Any developments in Tactical Doctrine which might affect subsequent designs of the BGT.

h. The Training and experience of BGHQ members.

i. The relative costs of the various means of this type of training.

RESULTS

This study was completed in June 1982 and a written report forwarded to the Director General of Army Training (UK) in September 1982. The study team were able to present a rigorous and objective evaluation of the present BGTs and make recommendations for the future development of command group simulators. By using the technique labelled "Scenario Analysis" a task list was produced with a level of detail that ensured effective communication between those involved in this study. The "Scenario Analysis" methodology is likely to be of value when considering the analysis and evaluation of complex command group operations. The production of collective training objectives using a "top down" approach may also be achieved using this technique.

AD P000841

IMPLEMENTING THE CRITERION REFERENCED USAF APPRENTICE KNOWLEDGE TEST PROGRAM

1Lt Christopher M. Antons
USAF Occupational Mesurement Center

An Air Force Apprentice Knowledge Test (AKT) is designed to measure specialty knowledge at the three-skill or apprentice level of a specific Air Force enlisted specialty. The Occupational Test Development Branch of the USAF Occupational Measurement Center (USAFOMC) is responsible for developing and maintaining the AKTs. AKTs are used in conjunction with other factors to select airmen for bypass of technical training and direct entry into a specific career field at the apprentice level.

Prior to 1976, AKTs were 65-item multiple choice tests with passing scores set annually at the thirtieth percentile of the score distribution for all examinees who had previously taken a specific AKT. Inherent to the method, passing scores fluctuated, sometimes dramatically, depending upon the examinee population for a given year. AKT use in some specialties was very low, thereby severely limiting the reliability of the passing score. Conversely, for high usage AKTs, a change of only one point for the calculated passing score on this relatively short test meant a large difference in the total number of airmen passing or failing. For the few 65-item AKTs still in existence, the passing scores range from 19 to 26 raw score points. The most severe limitation of this method was the fact that the passing scores were established relative to the examinee population without reference to job incumbents or expected performance.

To improve the AKTs, USAFOMC initiated a series of studies. In the first study, AKT scores were compared for three groups: beginning trainees and graduates of a technical training course for general vehicle maintenance, and airmen already selected for bypass in that specialty (Vaughan, 1976a). Mean scores for both the beginning trainees and bypass group were significantly lower than the mean score for graduates. Differences in scores of beginning trainees and graduates showed the test was able to discriminate among levels of knowledge for a specialty. Differences in scores of the bypass group and graduates demonstrated specialty knowledge differences between a group seeking apprentice skill level and a group just completing formal technical training. In comparison, the score at the tenth percentile of graduates was the same as the score at the seventy-fifth percentile of the bypass group. Using the score just above the tenth percentile as a passing score, some airmen previously selected as bypass specialists would not be qualified.

A second study replicated the first study on an additional five Air Force specialties and found similar results (Vaughan, 1976b). Based on the results of these studies, the USAFOMC implemented a criterion referenced testing program for AKTs using technical training graduates as the criterion group and the tenth percentile of that group as the passing score. The rationale given for originally setting the passing score above the tenth percentile was that extremely low scores are likely to contain considerable error (Lord and Novick, 1978). Conversely, a higher passing score was decided against since it might prevent acceptable performers from being selected to bypass training. Performance of technical school graduates and selected bypass specialists from one of the five specialties in the previous study were compared (Vaughan, 1978). Performance of the bypass specialists was shown to be equal to or slightly better than the technical school graduates. This evidence supported the decision not to set the passing score any higher than just above the tenth percentile.

In 1978, the USAFOMC began converting all AKTs to 100 items and criterion referencing those tests with a high usage (greater than 25 administrations per year), and a large enough criterion group of technical school graduates. All AKTs were expanded to

100 items to increase their reliability. The criterion referencing anchored the performance of bypass specialist candidates on the AKT to a known level of performance of technical school graduates. This allowed us to assume that successful bypass candidates had at least as much knowledge as the lower ten percent of technical school graduates for a given specialty.

Two main problems were encountered. First, we assumed that a few members of the examinee group would lack motivation for testing since they had just graduated, were preparing to depart for duty assignments, and were aware that the test had no impact on their own training. The USAFOMC explained the significance of this testing to training personnel and, in turn, the graduating trainees. This helped dispel the motivation problem. The second problem involved subject-matter experts (senior noncommissioned officers brought to the USAFOMC from working units in each specialty to provide input on content of the tests). They wanted to increase the difficulty of the tests to insure that bypass specialists would be knowledgeable. Test developers at the USAFOMC explained that increasing the difficulty of the test would also decrease the average score of the criterion group. With a lower mean criterion score, the passing score would be set lower. If set low enough, some examinees might achieve a passing score by chance alone.

Current Status

As of September 1982, of 270 specialties the following number of AKTs are available.

TYPE	NUMBER OF SPECIALTIES	AVERAGE USAGE
Criterion referenced	87	86
Noncriterion referenced	45	25
No test	138	--

The AKT program now includes both criterion and noncriterion referenced tests. All AKTs are criterion referenced unless annual usage is too low to justify the criterion referencing. For many specialties, no AKT is construrted because training is mandatory or other reasons specific to the individual specialties.

The following table provides information on usage of AKTs. Airmen take the examinations to bypass technical training when first entering the service (bypass) or when changing from one specialty to another (retraining), or to demonstrate apprentice level competency after a period of on-the-job training (upgrade).

AKT UTILIZATION
(CY 81)

USE	TOTAL TESTED	PERCENT PASS
Bypass	3517	68%
Retraining	1771	82%
Upgrade	1951	84%
Total	7299	75%

(Jan-Jun 82)

USE	TOTAL TESTED	PERCENT PASS
Bypass	934	59%
Retraining	1334	80%
Upgrade	962	83%

The following table provides information on the passing scores established for both the criterion and noncriterion referenced AKTs.

PASSING SCORE DISTRIBUTIONS

	N	Mean	SD	Range	%Passing
Criterion referenced	81	42.96	8.37	26-60	78%
Noncriterion referenced	29	42.28	6.20	30-56	77%
(65 item)Noncriterion referenced	6	24.50	2.74	19-26	85%

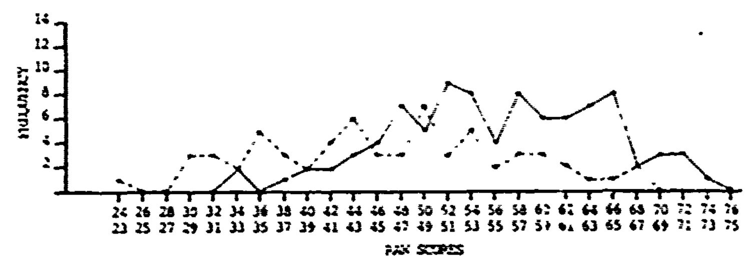
The average passing scores are nearly the same for criterion and noncriterion referenced tests. The difference, then, is not in placement of the passing score, but in the criterion that determines that score and the distribution of scores for that criterion group. As will be shown later, score distributions for the criterion groups are much less varied than for the examinee groups. Also, for % of examinees achieving passing scores, criterion and noncriterion referenced tests are nearly the same. According to the criteria for setting the passing score on noncriterion referenced tests, only 70% of examinees should pass. However, as stated earlier, the passing scores can fluctuate from year to year according to the population of examinees and the number of examinees passing depends upon the distribution of scores for one group compared to all past examinees. For the 65-item noncriterion referenced tests, there is more opportunity for fluctuation in scores from year to year and for examinees to achieve passing scores by chance.

Some Specific Criterion Referenced AKTs

Six criterion referenced AKTs were selected for analysis of both the criterion group and examinee group scores. Analyzing the AKTs individually, two were from specialties previously studied by Vaughan (1976a, 1978), two were selected for having extremely low passing rates and two were selected for having extremely high passing rates. Air Force Specialty Code (AFSC) 47230, Apprentice Base Vehicle Equipment Mechanic is similar to the general mechanic specialty examined by Vaughan (1976a). The passing score of 45 on this test is close to the average of 43 for all criterion referenced AKTs. In the 1976 study, the tenth percentile of the criterion group was the 75th percentile of the bypass group. In this case, the tenth percentile of the criterion group is the 47th percentile of the bypass group and the test is much more selective for the bypass group than the retraining group. For the purposes of the AKT, these characteristics are desireable.

47230
APPRENTICE BASE VEHICLE EQUIPMENT MECHANIC

	n	Mean	S.D.	10th Percentile
——— Criterion Group	91	55.56	9.07	44
- - - - Examinee Group	65	46.56	10.45	
Bypass	57	46.18	10.86	47% failed
Retrain	8	49.26	6.78	25% failed
Upgrade	0	-	-	-

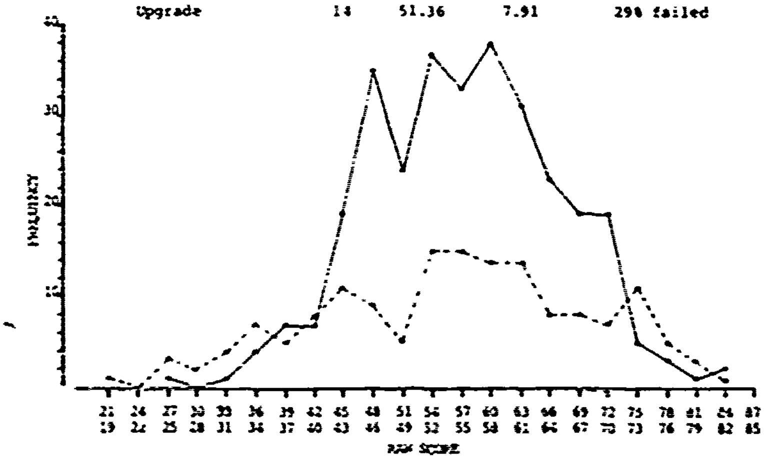


AFSC 90230, Apprentice Medical Service Specialist, was used in the performance measurement study (Vaughan, 1978) and criterion referencing study (Vaughan, 1976b). The passing score of 46 is also near the average for all criterion referenced AKTs. 33%

of the bypass group were below passing on this test compared to 58% in the 1976 study. Though means of examinee and criterion groups were nearly the same, the examinee group had the greater variance. The variance was not due to subgroups, since bypass and retrain groups both had large variance with their standard deviations twice the difference of their means. The upgrade group had less variance but was a smaller group and had a mean similar to the bypass group. What was notable was the large percent passing in the bypass group, indicating that, for this career field, civilian experience may provide adequate background. Considering the variance of the examinee groups, the cutoff scores were able to discriminate among examinees despite the similarity of examinee and criterion mean scores.

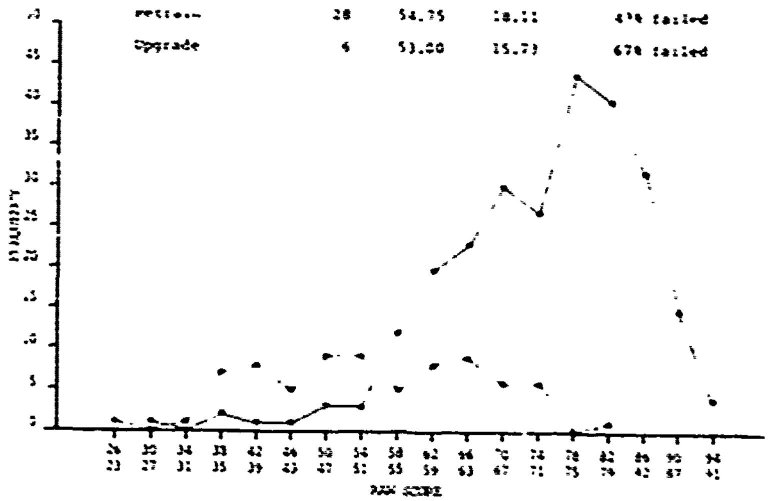
90230
APPRENTICE MEDICAL SERVICE SPECIALIST

	n	Mean	S.D.	10th Percentile
——— Criterion Group	309	56.34	9.72	45
- - - - - Examinee Group	170	54.41	14.69	
Bypass	79	51.28	14.63	33% failed
Retrain	77	58.18	14.94	20% failed
Upgrade	14	51.36	7.91	29% failed



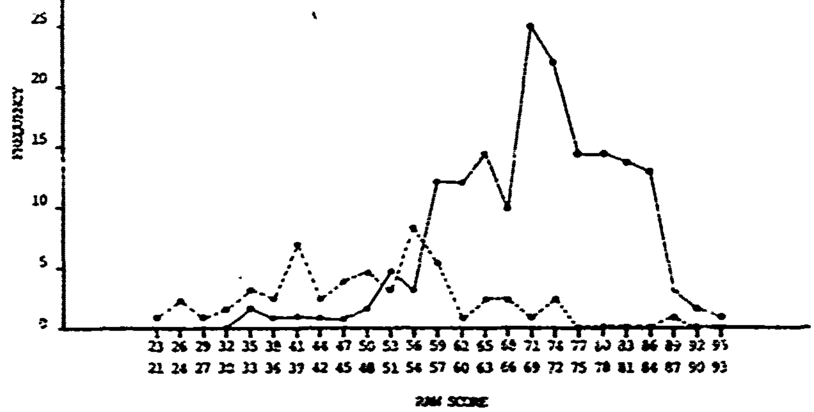
55235
APPRENTICE PLUMBING SPECIALIST

	n	Mean	S.D.	10th Percentile
——— Criterion Group	257	71.05	10.85	59
- - - - - Examinee Group	76	52.33	14.27	
Bypass	42	50.62	11.22	55% failed
Retrain	28	54.75	18.11	43% failed
Upgrade	6	53.00	15.73	67% failed



55330
APPRENTICE ENGINEERING ASSISTANT SPECIALIST

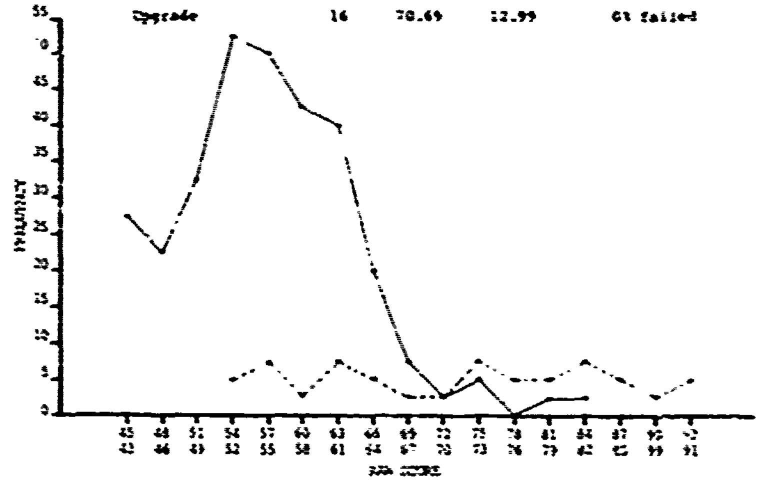
	n	Mean	S.D.	10th Percentile
——— Criterion Group	224	70.40	10.78	55
----- Examinee Group	71	49.27	13.21	
Bypass	53	43.09	11.63	88% failed
Retrain	30	53.53	12.90	53% failed



Two specialties, Plumbing and Engineering Assistant, showed high failure rates. Passing scores of 60 and 56 respectively were relatively high. Again, the examinee scores were highly varied. For the plumbing specialty, although the failure rate for the bypass group is the highest, the failure rate for the retraining group is also high. This suggests that those retraining may be coming from a variety of career fields and do not have the background knowledge required. For the engineering assistant specialty there is a much higher failure rate in the bypass group than in the retrain group. This suggests that knowledge required for this specialty may not be acquired in a civilian related job or there may not be a related civilian job. Again, the retrain group has a high failure rate that may suggest that those retraining are coming from a variety of career fields and lack the needed background knowledge. Also, the criterion group scores are higher than typical. Content of the test and training quality and emphasis may contribute to this effect.

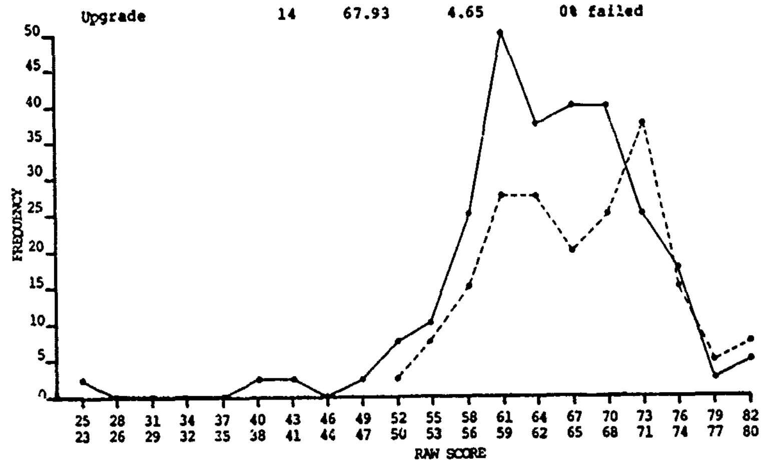
20630
APPRENTICE IMAGERY INTERPRETER SPECIALIST

	n	Mean	S.D.	10th Percentile
——— Criterion Group	123	55.97	7.59	45
----- Examinee Group	28	71.15	12.75	
Bypass	1	50.00	-	0% failed
Retrain	11	73.73	16.06	0% failed
Upgrade	16	70.69	12.99	6% failed



24130
APPRENTICE SAFETY SPECIALIST

	n	Mean	S.D.	10th Percentile
——— Criterion Group	108	63.79	8.23	56
----- Examinee Group	78	65.23	11.85	
Bypass	0	-	-	-
Retrain	64	64.64	12.85	9% failed
Upgrade	14	67.93	4.65	0% failed



AFSCs 20630, Apprentice Imagery Interpreter and 24130, Apprentice Safety Specialist, had AKTs with few or no failures. Both were different from the other AKTs studied in that only one test was administered for bypass and most tests were administered to Air Force Reserve and National Guard members either for retraining or upgrading purposes. Higher examinee means would be expected for these groups than for bypass groups. Airmen taking these exams may have already worked in the specialty or a very closely related specialty. In the case of the Safety specialty, some knowledge of that field is required for all specialties.

In general, all six AKTs exhibited two distinct characteristics. First, the variance in the distribution of scores was always greater for the examinee group than the criterion group. Though it can be expected that the criterion group, having just completed training in a specialty, would not vary much on a test covering that specialty, it was somewhat less expected that the examinee group scores vary so much more than the criterion group. For the Apprentice Medical Service Specialist test, the standard deviation for the examinee group was nearly five points greater than for the criterion group. Second, in looking at the subgroups of examinees, those taking the test for retraining and those for upgrading always had higher mean scores than those in basic training trying to bypass technical training. This result can be expected since those airmen retraining and testing for upgrading have been in the Air Force for a period of time already and have had an opportunity for more specific experience or, in the case of testing for upgrading, have been through on-the-job training in the specialty. These characteristics indicate that the criterion referenced tests are able to discriminate across varied groups of examinees.

Conclusions

For the AKTs analyzed, the higher means and relatively small standard deviations of the criterion groups provide a more precise pass/fail cutoff. It can be seen from the score distributions that when the scores at and below the tenth percentile for the criterion group are eliminated, the criterion group has greater homogeneity so that selection for bypass or retraining is similar to membership versus nonmembership in the criterion

group rather than achieving a specified criterion percentile across a distribution of criterion scores. This serves the expressed purpose of the AKTs to provide a means of selecting or not selecting an individual to bypass technical training.

For those AKTs with either very low or very high passing rates, criterion referenced tests were able to discriminate where the noncriterion referenced tests would have allowed too many or too few passing scores respectively.

Recommendations

Given large differences in mean scores of examinee and criterion groups, it is difficult to determine the validity of very high or very low passing rates. Performance studies of the bypass groups (Vaughan, 1978) should provide validity for the criterion cutoff scores. We are directing future research toward this goal.

Additionally, the high variance of examinee groups analyzed indicates a need for screening potential examinees. Given the wide range of examinee scores, some tests may be administered to airmen lacking the appropriate background knowledge or experience needed for a specialty. This suggests overuse of the tests and need for a better screening procedure.

References

Lord, F.M., and Novick, M.R. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.

Vaughan, D.S. Criterion-referencing the 47330 Apprentice Knowledge Test to technical school graduates. Lackland Air Force Base, Tx.: USAF Occupational Measurement Center Technical Note Series, March 1976. (No. 76-01)

Vaughan, D.S. Development of criterion-referenced Apprentice Knowledge Tests. Lackland Air Force Base, Tx.: USAF Occupational Measurement Center Technical Note Series, May 1976. (No. 76-04)

Vaughan, D.S. Job performance of USAF bypassed specialists. Oklahoma City, Oklahoma: proceedings of the 20th Annual Conference of the Military Testing Association, 1978.

Acknowledgements

The author gratefully acknowledges the following people for their assistance on this project: Mr. Paul Stanley, Chief, Occupational Test Development Branch; Lt Col Jimmy Mitchell; Maj Martin Dittmar; Capt Robert Worley; Ms. Marian Dike; Mr. Jim Barlow; SSgt Denise Born; Sra Anthony Jackson; A1C Bryan Hassenpflug; Mrs. Irma Bozart; Mrs. Velma Grantham; Ms. Anne O'Neill; and all the word processing section.

Performance Appraisers as Test Developers

Cristina G. Banks, Ph.D.
Department of Management
The University of Texas at Austin

Numerous authors have expressed disappointment in the lack of success organizations experience with most performance appraisal systems (Carroll & Schneier, 1982; Landy & Farr, 1980; McCall & DeVries, 1977). Most complaints center on the plethora of intentional and inadvertant biases that permeate performance ratings (Landy & Farr, 1980) and the lack of acceptance of performance appraisal systems by users (DeVries, Morrison, Shullman, & Gerlach, 1981). Numerous strategies were implemented to improve the psychometric qualities of performance ratings, the most common consisting of more effective rater training and improved format development. We now know that these strategies have reaped few positive results (Dunnette & Borman, 1979; Landy & Farr, 1980). Despite improvements in the mechanics of appraisal (e.g., format clarity, increasing appraisers' awareness of the importance of appraisal), ratings are still fraught with errors and management still has a difficult time getting people to complete appraisals Start accurately. Our current research emphasis on the appraisal process and the impact of organizational context variables may reach the same end. All past and current efforts to improve the quality of performance appraisals may be doomed to failure because they fail to take into account the demands of the appraisal task on the appraiser. That is, the appraisal process requires appraisers to be good test developers.

While at first this appears to be a strange idea, a closer examination of the rater's role in the appraisal process will reveal the link between appraisal and test deveopment. First, consider the activities a rater undergoes when engaged in a rating task. The rater observes an employee performing job-related duties and records (in memory or on paper) instances of job-related behavior that could be used to evaluate specific dimensions of performance. At an appropriate time, the rater recalls recorded employee information and through some process relates this information to the criteria provided in the appraisal format to generate summary performance evaluations for each dimension.

How does the appraisal format assist the rater in this task? For illustration, we will use behaviorally-anchored rating scales (BARS) to answer this question since BARS have been characterized as the preferred type of rating format (Carrol & Schneier, 1982) and enjoy widespread use. Other formats such as graphic rating scales and mixed standard scales could be examined in a similar way. The question we must pose is, what do BARS present to the rater to facilitate his or her assessment of employee performance?

In most cases, BARS consist of a set of scales, each describing one important aspect of job performance, called performance dimensions. Each dimension becomes the title of the scale accompanied by a single continuum marked off in units representing scale points. Scale points are anchored by one or more specific examples of employee behavior. These anchors are selected on the

basis of their ability to serve as unambiguous indicators of specific levels of performance along the continuum. Raters scan the anchors for each performance dimension and select the anchor that best reflects a ratee's typical job behavior (Carroll & Schneier, 1982, p. 112). The point value associated with the anchor becomes the performance rating for that dimension. Thus, from the user's point of view, BARS provide separate judgments for each aspect of job performance and with each dimension, provide examples of specific job-related behavior for the rater to "scale" recalled instances of the ratee's behavior on the job. The advantages of BARS include a focus on actual job behavior (and not vague trait or global characteristics), an ability to specify exactly what an employee needs to do to receive high ratings, allowing raters to give feedback and specify why employees received the ratings they did, and providing greater awareness of important aspects of job performance (Carroll & Schneier, 1982, pp. 112-113). Given this description of BARS, do they actually aid the appraisal process?

Consider, again, what a rater does when evaluating an employee with a BARS format. If a rater participates in scale development, he or she is exposed to considerable discussion of the differentiation of dimensions and more importantly, the specific job behaviors which illustrate different levels of performance. This essentially "teaches" the rater what is relevant and irrelevant to observe on the job. In addition, persons who participate in scale development are taught approximately the same rules by which they selectively observe and evaluate ratee behavior. When the rater does not participate, he or she depends on one's own experience and previous training regarding relevant and irrelevant ratee behavior. Regardless of the degree of involvement, raters selectively observe and evaluate ratee behavior and later recall this information when completing an appraisal. When using a BARS format, the rater must review recalled information and make a summary judgment of the level of performance indicated by the behavior before the anchors on the format can be utilized. This is because one must make a judgment concerning a person's typical performance level on the job before one can make a judgment of the similarity between an anchor and the ratee's typical performance. The only way one can judge an expectation that an employee will behave in a particular manner is to have a standard with which to compare it to, and in this case, the standard is a summary evaluation of the types of behaviors an employee has or will engage in. By this point, the appraisal process has already occurred; that is, relevant items (job-related behavior) have already been selected and evaluated along some criterion (which is probably schema driven) before anchors play a part in the appraisal process. The anchors provide a way of setting a numerical value to the summary judgment by relating the probability an employee will engage in a specific behavior to the cluster of behaviors the employee has engaged in previously. And it might be the case that when the numerical value associated with the preferred anchor does not correpsond to the summary evaluation previously formed, anchors may be ignored in favor of matching scale values to the previously formed judgment. Anchors would be ignored to the extent the rater does not share the same criteria for evaluating job-related behaviors with the scale developers. Thus, it can be argued, that the BARS format has

little impact on the appraisal process itself, and merely allows a scale point to be assigned to judgments already formed on the basis of a rater's idiosyncratic criteria.

Why doesn't the BARS format have a greater impact on the appraisal process? The answer lies in the lack of sound psychometric properties built into an appraisal format as a test. After a thorough review of well-known texts on test construction (e.g., Anastasi, 1981; Cronbach, 1970, Ghiselli, Campbell & Zedeck, 1981; Guion, 1965), several properties of tests were identified as critical for the reliability and validity of assessments. Those particularly relevant to performance appraisals are listed in Figure 1. These may be grouped into four general areas: selection of items (A), item power (B), scoring and interpretation of scores (C), and contaminating effects (D). With respect to the first group, the general problem is that BARS do not provide many items the rater can use to evaluate employee performance. The only items given are the anchors which are few in number, are not intended to representatively sample the behavior domain, and are developed on the basis of experts' notions of performance schema which may or may not be similar to the rater's schema. The lack of items on which to judge performance forces raters to develop their own set of items (e.g., job-related incidents) of unknown intercorrelation, representativeness, discriminability, validity and bias. The power of these items is unknown because these items are never explicitly stated and hence, not empirically tested, leaving little opportunity for feedback necessary for item revision. Also, when these items are maintained within the rater's mind, there is no opportunity to develop schemes for optimally weighting and combining items. Whatever strategy a rater formulates to select and score items, the problem of interpreting scores in a meaningful way also must be overcome. The rater does not have at his or her disposal a table of norms from which to interpret scores and thus, must relate scores to previous scores obtained -- hardly an unbiased standard and one that can be generalized to employees in general. When items are contained within the minds of raters, there is little opportunity to standardize procedures (approaches, strategies, etc.) and to ensure assessment reliability. Thus, we can expect unstandardized item development and scoring of items with unknown discriminating power and usefulness, and idiosyncratic interpretation of scores that may differ not only across raters but also across ratees for the same rater.

Several contaminants also contribute to the lack of reliability and validity in appraisals. Because performance judgments are so dependent on the input of the rater, appraisals become highly susceptable to factors which may alter the rater's ability to conduct an appraisal consistently such as the rater's emotional state (e.g., fatigue, motivation) and skill (e.g., knowledge of appraisal, personal bias). Obviously, if the scoring and interpretation of items occur within a rater's head, the appraisal is susceptible to "subjective" as opposed to "objective" scoring, contributing further to appraisal unreliability. In addition, in the process of searching for relevant ratee information, it is possible that early information collected may color one's evaluation of subsequent information, resulting in an assessment that reflects a response set more than an

employee's status on a performance continuum.

It is evident by examining appraisals from a test construction view that appraisal formats offer little direction for properly evaluating employee work performance. At least in the case of BARS, the format merely defines the construct (i.e., dimension) to be rated and offers illustrative items. The rater must develop a strategy for identifying, scoring and weighting relevant items (i.e., job-related incidents) and then interpreting employee scores. In essence, each rater develops his or her own "test" of job performance. The actual use of the BARS format comes at the end of the appraisal process; that is, when a numerical value is assigned to the employee's score on the rater's own "test." Thus, the appraisal format is merely incidental to the appraisal process, severly limiting its impact on appraisal success. Therefore, what is likely to be a major determinant of rating quality is the appraiser's (e.g., rater's) skill in developing valid performance tests, not appraisers' ability to use the format properly.

So, what does this perspective of performance appraisal as test development say about efforts to improve performance ratings? This paper has discussed the tasks a rater must engage in, and shown that the formats typically used fail to aid the rater in completing those tasks. Yet the continual problems of low validity and reliability of ratings indicate that raters are unable to complete the tasks themselves -- that is, they are not naturally good test developers. I believe that efforts to improve performance ratings will need to address these two flaws. Either the rater must be taught how to independently develop and use a valid test, or a rating format/task must be constructed that provides raters with the aids they need.

The view of rating as test development can be useful in investigating both strategies. For example, what kind of rating format is likely to result in improved ratings? The checklist shown in Figure 1 provides some clues. It will need to have valid items readily available, so that all raters utilize the same ones, and have optimal scoring systems already defined. A method that might possibly meet these criteria is job sampling, where raters need only observe behaviors that have or have not occurred, and idiosyncratic judgments and inferences are eliminated. Behavior assessment (Cone, 1980; Komaki, Collins & Thoene, 1980) may be another viable option. Again this method utilizes the rater only as an observer and recorder of events.

The design of rater training programs can also be facilitated using a test development perspective. In training, the objective would be to teach raters how to construct valid tests and to insure that all raters use similar tests. Again, a glance at Figure 1 shows the various behaviors that must be taught in a training program. Raters must learn to select appropriate behaviors, evaluate them in a certain way, and combine their evaluations into a summary judgment. This implies that training must change and standardize raters' cognitive structures and strategies, which has been suggested by other writers (Cooper, 1981; Feldman, 1981). However, previous writings on cognitive processes in performance appraisal have failed to provide clear objectives for training design. By breaking up the rating process into the steps of test development, the behaviors needed to perform the rating task are delineated and training methods

designed explicitly to change those behaviors can be developed.

In conclusion, the intervention strategies currently available to improve performance ratings (e.g., formats, participation, training) have no theoretical base. Not surprisingly they have come to dead ends. Although cognitive theories of rating are gaining popularity , they are not developed to the point where specific interventions can be suggested. The test development perspective explicitly ties the rater's cognitive processes to the appraisal context, and thus points to specific interventions that also have theoretical potential for improving rating quality.

References

Anastasi, A. Psychological testing (5th ed.). New York: Macmillan, 1981.

Carroll, S. J. & Schneier, C. E. Performance appraisal and review systems. Glenview, Illinois: Scott Foresman, 1982.

Cone, J. D. The overlapping worlds of behavioral assessment and performance appraisal. Paper presented at the First Annual Scientist-Practioner Conference in Industrial/Organizational Psychology, Old Dominion University, Virginia, 1980.

Cooper, W. Ubiquitous halo. Psychological Bulletin, 1981, 90, 218-244.

Cronbach, L. J. Fundamentals of psychological testing. New York: Harper & Row, 1970.

DeVries, D., Morison, A., Shullman, S., & Gerlach, M. Performance appraisal on the line. New York: Wiley-Interscience, 1981.

Dunnette, M. D. & Borman, W. C. Personnel selection and classification systems. Annual Review of Psychology, 1979, 30, 477-525.

Feldman, J. Beyond attribution theory: Cognitive processes in performance aprpaisal. Journal of Applied Psychology, 1981, 66, 127-148.

Ghiselli, E. Campbell, J., & Zedeck, S. Measurement theory for the behavioral sciences. San Francisco: W. H. Freeman, 1981.

Guion, R. Personnel testing. New York: McGraw-Hill, 1965.

Komaki, J., Collins, R. L., & Thoene, T. J. F. Behavioral measurement in business, industry and government. Behavioral Assessment, 1980; 2, 103-123.

Landy, F. & Farr, J. Performance rating. Psychological Bulletin, 1980, 87, 72-107.

McCall, M. W. & DeVries, K. L. Appraisal in context: Clashing with organizational realties. Technical Report #4, Greensboro: Center for Creative Leadership, 1977.

Figure 1

Quality of 'Good' Tests	Definition	Observed in PA's
(A) Representative sampling of behavior domain	The extent to which items cover the range of behaviors included within the construct	No
(A) Multiple items	The degree to which construct is measured by several items	?
(A) Interitem correlation	Internal consistency of items	No empirical test
(A) Construct well defined	The degree to which it is clear what the construct does and does not cover	?
(A) Items developed from input of several experts	The degree to which items do not reflect a single idiosyncratic view	No
(B) Item discriminability	The degree to which the item correctly separates people into high and low criterion groups	No empirical test
(B) Item validity	The degree to which item predicts meaningful criterion	No empirical test
(B) Optimal weighting of items	The manner in which items may be combined to maximize prediction of a meaningful criterion	No empirical test
(B) Items pretested in representative, heterogeneous sample	The degree to which items are generalizable to other samples	No
(B) Unbiased, 'culture-fair' items	The extent to which items do not inherently favor particular subgroups	No empirical test
(C) Standardized scoring procedure	A standard approach to scoring items	No
(C) Norms available for score interpretation	A standard available to which a single score may be compared	No
(C) Mechanism for checking reliability of scoring	Ability to check for scoring errors	No
(C) Uniformity of test procedures	Consistency in the application of test procedures	No
Contaminants		
(D) Items "trigger" a response set	The probability that one item will lead to a closed view of responses from following items	Likely
(D) Susceptibility to examiner's emotional states (anxiety, motivation)	The probability emotional factors affect item selection, scoring and interpretation	Yes
(D) Susceptibility to subjective scoring	The degree to which scoring criteria are unobservable and idiosyncratic	Yes
(D) Susceptibility to examiner's idiosyncracies (skills, bias, contamination)	The degree to which item selection, scoring & interpretation dependent on examiner characteristics	Yes

VALIDATION OF THE IRJ PROCEDURE FOR ANALYZING THE RATING PROCESS

Cristina G. Banks, Ph.D.
Department of Management
The University of Texas at Austin

Jody L. Toquam
Personnel Decision Research Institute
Minneapolis, Minnesota

What seems to be a new trend in industrial/organizational psychology is to examine the decision-making processes in performance appraisal. It appears to be a fad in the sense that only very recently researchers have seriously considered "process" issues, and despite the recency of this awakening, almost everyone seems to be doing it. This is evidenced by the proliferation of process-oriented symposia at the latest American Psychological Association meeting (1982). Like most fads, the value of this approach has been accepted uncritically, and like most participants in fads, researchers are jumping into studies with paradigms borrowed from other areas and disciplines (e.g., social cognition, policy capturing) without examining their fit to the unique specifications of the appraisal task. This paper addresses the issue of paradigm fit by testing the validity of a specific paradigm and assessing its value for gaining insight into rating accuracy.

In presvious research, the author developed a new paradigm for studying the appraisal process that was believed to fit reasonably well with the kinds of questions that need to be answered in the appraisal context (Banks, 1979; 1980). Validity data were not available at that time to support this approach though preliminary analyses supported its potential (Banks, 1982). This paper presents three types of evidence which lend strong support for the value of this paradigm and for the importance of process approach in general. First, the paradigm is briefly described and then the three sources of validity evidence are presented.

Instantaneous Report of Judgments (IRJ)

Instantaneous Report of Judgments or IRJ was developed by the author to capture some of a rater's cognitive processes during an appraisal task. Briefly, a rater reports his or her judgments formed during a rating task by using a panel of buttons to record judgments of ratee performance and by reporting verbally behavioral cues that "trigger" judgments (see Banks, 1980 and 1981 for more detail). Basically, IRJ provides raters a mechanism for reporting the contents of their decision-making whenever they feel the "urge" to report.

A typical IRJ task is to present a videotaped performance of a manager in a job and ask the rater to evaluate the manager's performance along a single performance dimension (e.g., "Establishing and Maintaining Rapport"). The rater is instructed to press a button whenever one "feels" he or she is making a judgment, and to press the button (1 to 7; 1 = low effectiveness, 7 = high effectiveness) that best represents his or her judgment of ratee performance. After pressing a button, the rater reports verbally the basis for his or her judgment. Raters are encouraged to press buttons as many times as they make judgments, and at the conclusion of each task, the rater renders a summary rating. Typically, a rater views and rates several ratees in order to obtain multiple samples of raters' cognitive

processing.

Several behavioral indices of raters' cognitive processing are obtained: (1) number of judgments made per task; (2) variation in judgments within tasks; (3) variation in mean judgments across tasks; and (4) latency before initial judgment. These indices are operationalizations of four constructs believed to be related to rating ability: (1) degree of information utilization; (2) sensiti ity to ratee differences; (3) sensitivity to ratee strengths and weaknesses; and (4) global vs. specific observational style, respectively. In addition, it is possible to trace the specific information utilized by a rater during a task. These behavioral indices of the rating process provide access to rater behavior while judgments are formed and thus, allow the possibility for gaining insight into the determinants of rating accuracy. This research focuses on the segment of performance appraisal that entails initial formation of judgments from observation only; it does not shed light on the recall of information nor on the factors that affect how stored evaluations are combined to generate final appraisal ratings.

Three studies were conducted to evaluate the meaningfulness of these data. Study 1 investigated the impact of reporting on the appraisal process. It sought to determine whether reporting the contents of raters' thought processes disturbs the rating process, hence limiting the generalizability of these studies. Study 2 examined the relationship between the cues (or information) raters select and rating effectiveness. This study sought to determine whether more accurate raters tended to select different kinds or amounts of information than less accurate raters. Finally, Study 3 examined the relationship between various indices of rater behavior and rating accuracy. All three studies were undertaken to test the construct validity of the IRJ procedure and thus provide some measure of the meaningfulness of the rating behavior measured.

Study 1

To examine the impact of reporting judgments on the rating process, overall ratings obtained from IRJ tasks were compared with those obtained from typical rating tasks. Two samples of raters participating in IRJ studies were compared with two independent samples, one from Borman's rating reliability and accuracy study (Borman, 1979) and one collected recently by the author. The two non-IRJ studies consisted of raters viewing the same managerial performances as in IRJ tasks and simply recording overall performance ratings. Mean performance ratings per task (managerial performance) were calculated for each of the four samples, and these are shown in Table 1. Mean ratings were correlated between IRJ and non-IRJ samples to determine their similarity in rating outcomes. As can be seen in Table 2, the mean ratings are highly correlated between both types of studies (r's at least .90; $p < .01$). This suggests that despite differences in procedure, samples, and rating instructions, rating outcomes were remarkably similar. When differences between pairs of mean ratings are examined, almost all differ by less than one scale point and the sum of the differences are near zero (e.g., Borman & IRJ1 sample = .3). It appears from these data that the reporting requirement for subjects participating in IRJ studies does not distort the rating process and thus, we can say that findings from

IRJ studies are probably generalizable to more typical rating studies.

Table 1

MEAN PERFORMANCE RATINGS FOR 20 RATING TASKS FOR IRJ AND NON-IRJ SAMPLES

Tasks	IRJ1	IRJ2	Borman	Banks
1	1.95	2.23	3.18	2.61
2	4.81	4.51	4.00	4.12
3	4.03	3.14	3.72	3.48
4	4.08	3.17	3.99	3.69
5	3.57	2.58	3.73	3.81
6	4.08	3.21	3.64	3.71
7	6.56	6.45	5.64	6.71
8	5.29	4.92	3.91	4.46
9	5.88	5.68	4.66	5.33
10	5.00	5.50	4.50	4.94
11	4.92	4.98	3.49	5.30
12	4.59	4.18	4.12	5.17
13	3.57	3.14	3.63	3.17
14	1.85	1.46	2.88	2.15
15	2.58	2.28	3.47	2.74
16	2.56	1.82	2.88	2.15
17	2.12	1.82	3.30	2.61
18	1.81	1.53	3.22	1.87
19	3.08	2.58	3.08	3.02
20	1.95	1.75	1.77	2.35

Table 2

Intercorrelations of Mean Performance Ratings Between Four Independent Samples

	IRJ 1	IRJ 2	Borman	Banks
IRJ 2	.97			
Borman	.90	.91		
Banks	.94	.96	.91	XXX

Study 2

As part of a larger study of cue selection and evaluation, the author examined the types of cues raters utilize in their judgments. Raters were divided into high and low deviation groups on the basis of the absolute difference between their rating and the corresponding expert rating for the same managerial performance. Raters were divided into high and low deviation groups by median split (excluding the middlemost score). Basically, raters who deviated greatly from the expert rating were "high deviators" and those who deviated little from the expert rating were "low deviators." While deviation scores are only crude estimates of accuracy, they may be sufficient to gain insight into cue utilization.

Behavioral cues reported verbally by raters were examined for each group for two ratees (A and C). The frequency of report for each behavioral cue were determined for each of six performance dimensions per ratee. Then the authors classified cues as either "relevant" or "irrelevant" for evaluating each performance dimension without knowledge of cue frequencies within groups. The frequency of report of relevant and irrelevant cues are reported in Table 3. The table shows that irrelevant cues are reported with a much higher frequency by high deviators than by low deviators, and for some dimensions, deviators are further differentiated by a higher report frequency of relevant cues by low deviators than by high deviators. A test of the significance of a Deviation X Category (relevant vs. irrelevant) interaction obtained in a multilevel contingency table analysis was significant for both ratees (likelihood x = 99.7 & 181.73 $p < .01$). Thus, raters who provide ratings most similar to those of experts select and evaluate different sets of cues from those of less similar raters. This suggests that IRJ is sensitive to identify cues utilized by raters and more important, can capture important differences in cue selection strategies. This evidence provides considerable support for the internal validity of the IRJ procedure.

Study 3

In this study, various indices of rater behavior elicited during the rating process were correlated with accuracy, halo, leniency, and restriction of range. Accuracy, halo, leniency, and restriction of range were calculated according to accepted conventions (see Borman, 1979, for details). Correlations were calculated separately for a student sample ($N = 23$) and a manager sample ($N = 33$). Correlations between four rating behaviors — judgment frequency (AVGNJ), variation in judgments (AVGSDJ), variation in mean judgments (AVGSD), and latency of first judgment (AVGLAT) and the four rating outcomes are shown in Table 4. It appears from the table that restriction of range error is related to AVGSD as it should be since AVGSD is the micro-level analog of restriction of range; both measure differentiation between ratees. This correlation suggests that differentiation (or absence of it) at the judgment level is consistent with differentiation at the summary rating level. For managers, leniency was related to the number of judgments a rater made (AVGNJ) and the variation in judgments (AVGSDJ), suggesting that

Table 3

Frequency of Relevant and Irrelevant Cues

Category		Dimensions					
Ratee A							
REL	H	39	21	19	22	25	19
	L	55	30	15	28	18	13
IRREL	H	47	43	64	67	39	50
	L	33	41	51	53	21	55
Ratee C							
REL	H	35	58	19	17	23	8
	L	19	72	21	7	26	11
IRREL	H	60	43	79	98	99	42
	L	31	69	63	39	38	56

the more judgments a rater makes and the greater the differentiation in judgments per ratee, the lower the leniency. These correlations, however, were found only in the manager sample.

Of greater importance are relationships with accuracy For students, accuracy was related to the judgment frequency, variation in judgments, and latency; that is, accurate raters tended to make fewer judgments, exhibit less variation in judgments, and take more time generating the first judgment than less accurate raters. Conversely, no significant correlations were found for the manager sample.

Table 4

Correlations Between Rating Behaviors and Rating Outcomes for Managers and Students

	MANAGERS			
Rating Behaviors	Rating Outcomes			
	Halo	Leniency	Rest. Range	Accuracy
AVGNJ	25	-.35**	.18	.01
AVGSDJ	.15	-. 5	0	-.01
AVGLAT	.09	[illegible]	-.15	-.12
AVGSD	.12	.37**	.84***	-.03
	STUDENTS			
AVGNJ	-.11	.10	.22	-.53**
AVGSDJ	-.05	.36	-.22	-.43**
AVGLAT	.23	-.13	.04	.46**
AVGSD	.13	.01	.66***	.13

*p<.05
**p<.01
***p<.001

Results observed in the student sample could explain those found in Study 2; that is, accurate raters make fewer judgments because they select fewer irrelevant cues than less accurate raters. Thus for student raters, a more conservative reporting style seems to be more effective. Low judgment frequencies were also associated with smaller variations in judgment and longer latencies. These observations reinforce the notion that student raters were probably processing ratee information at a more global or abstract level.

The lack of significant correlation in the manager sample was puzzling at first. However, a scatterplot of the relationship between judgment frequency and rating accuracy revealed a moderate curvilinear relationship (eta = .39), suggesting that for managers, two styles of rating behavior may be adaptive: one which is similar to the students' style, and the other which utilizes and processes specific ratee information. The latter style could be developed over time with growing knowledge of the appropriate interpretation of ratee behavior through job and appraisal experience. Highly experienced managers would have then the necessary cognitive "schema" to interpret subtle behavior cues, resulting in high judgment

frequency, large variation in judgments, and short latency. At this point, however, this is purely speculative.

Results from Study 2 and Study 3 suggest what some of the linkages between rating behavior and rating accuracy might be. Clearly, for students, a more conservative, general processing style results in greater accuracy, but this does not hold true uniformly for managers. For managers, cue selection and evaluation strategy may be dependent on the extent of a manager's experience in the job, familiarity with appraisal issues, and knowledge of the specific content of the managerial job.

Conclusion

Results observed in these three studies suggest that Instantaneous Report of Judgments permits one to: (1) obtain detailed process information without disturbing or altering the judgment process and ratings that result; and (2) specify how more accurate raters differ in some rspects from less accurate raters. This paper provides only a brief look into the kinds of rating process information IRJ can provide. Potentially, the IRJ procedure can provide a wealth of useful information. This basic information about process is necessary for organizations to design potent rater training programs and appropriate appraisal formats to maximize accuracy. Indeed, process research need not be a fad.

References

Banks, C. G. A laboratory study of the decision-making processes in performance evaluation. Unpublished doctoral dessertation, University of Minnesota, 1979.

Banks, C. G. IRJ: A new technique for measuring the performance rating process. Proceedings of the Military Testing Association, Toronto, 1980.

Banks, C. G. Instantaneous Report of Judgments: A technique for analyzing the performance rating process. Department of Management Working Paper 81/82-4-6, University of Texas, 1981.

Banks, C. G. Behavioral indices of raters' cognitive processing in performance appraisal. Paper presented at the annual meeting of the America in Psychological Association, Washington, D.C., 1982.

Borman, W. C. Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 1979, 64, 410-421.

PROBLEM SOLVING AND STUDENT MOTIVATION FOR CLASSROOM INSTRUCTORS

by
G. M. Barry, Ed.D.
Educational Research and Development Center
The University of West Florida

Introduction

What is it that instructors do when they help students? Further, how is it that some instructors seem to have a special ability to help a student overcome a personal or academic problem that is a deterrent to that student's ability to learn in the classroom?

At the Educational Research and Development Center (ERDC) we have found that the following knowledge and skill areas are important parts of this process:

Skill Area	Behavior Indicators
1. Organization	Being able to be in control of a problem solving process with a student. This control can be natural, learned, or a combination.
2. Instructor styles and preferences	Being able to share with others their organizational and instructional strategies that demonstrate knowledge of themselves and their instructing style.
3. Student individual differences	Being able to recognize individual differences in students, knowing how to motivate students through their preferences, and giving them confidence in approaching the unknown.
4. Listening	Being able to recognize the level of listening that a student is attending when listening to the instructor's thoughts and ideas.
5. Problem solving	Being able to help the student identify problems without the instructor becoming solution oriented too soon.
6. Helping	Being able to use the helping relationship while working on problem identification.
7. Giving and receiving feedback	Being able to give and receive feedback while not creating an imbalance in the relationship.
8. Personal model of instruction	Being able to apply a strategy for working with individual students as well as groups that allow students to develop new insights into reasons for their problems and how they might improve.

Over the past two years, ERDC has designed and implemented a training program for military instructors for learning in the eight skill areas and perhaps more important, to try out these skills with real students. The training takes two and one-half days and has the following as its major goal:

GOAL: *To provide specific skills to enable instructors to help the student learner solve a variety of problems, many of them non-academic, which are deterrents to his/her success in a learning situation.*

For an educational research and development center to say that we came up with these eight areas by a careful needs assessment procedure, literature review, and a program design would probably be typical. Unfortunately, or fortunately, that was not the case. What did happen was that the program design was tied to the mission of our service oriented research and development center. We are charged by the legislature in Florida to deliver research-based workable programs to a seven-county area. In the process of experiencing and experimenting with training materials such as the Myers Briggs Type Indicator, (MBTI), (Myers, 1962) process training packages from the Northwest Regional Laboratory (Jung, Pino, Emery, 1972), and one of Joyce and Weil's Models of Teaching (1980), we came up with what has turned out to be a highly successful training sequence.

Even though we thought our research base was good, we have learned the importance of organizing, sequencing, and allowing for an actual tryout of the skills. Maximum program potential and post training evaluation improvements came when we began to do more live demonstrations and when we brought in students with real academic problems as part of a practicum experience.

This program is designed for one school organization at a time. Some educators would call this a school-based program. There are three important ingredients: (1) a high quality 2½ day training program for instructors; (2) an organizational phase that includes counselors and administrators; (3) a training program for students.

What is the Program Like?

The program is designed to put instructors more in control of instructor-student interactions. If instructors have skills available and a workable format to use the skills, they will be better able to help students with problems and motivate students toward specific learning goals. No attempt is made to make counselors out of instructors, but it is clear that there are not enough counselors to help all the students who need help. What is done in this program is to develop the personal aspects of instruction, especially problem solving and student motivation, that is usually carried out on a one instructor to one student basis.

Many of the skills are familiar to instructors and the training is a chance to tune up these skills. The context of individual differences is developed through the use of the Myers Briggs Type Indicator (MBTI). Other instruments which clarify perception and point out individual differences among students and instructors may work just as well, but the MBTI has the advantage in that it says positive things about individuals which is an important base for skill development for the rest of the training.

A major emphasis during the training program for instructors is the process of giving and receiving feedback. Since giving feedback can create imbalances in relationships, instructors often avoid giving feedback to students. This is an understandable situation given the lack of training and the real perception that instructors teach groups and not individuals. For the student who is different in their learning preference from the instructor, a dilemma develops. Now there is need for feedback if for no other reason than to set common learning goals. It is for this reason and others like it that a personal model of teaching is useful and practical.

Content Areas and Schedule

Instructor Training Phase

Major Content Areas

1. Personality Types and Preferences
2. Problem Identification
3. Helping Relationship
4. Learning and Teaching Styles
5. Giving and Receiving Feedback
6. Intervention Techniques and Demonstration
7. Try Out (Practicum)

Training Schedule

First Day: Course Objectives
Administration of the Myers Briggs Type Indicator
Introduction to Type
Feedback on the Myers Briggs Type Indicator
Case Studies
Learning and Instructor Styles
Myers Briggs Type Indicator Research

Second Day: Levels of Listening
Problem Identification and the Helping Relationship
Giving and Receiving Feedback
Academic Problem Solving Model and Demonstration

Third Day: Prepare for Practicum and Review
Student/Instructor Problem Solving
Feedback to Instructors
Student Interviews
Debrief Problem Solving Sessions
Complete Course Evaluation

Where the Learning Takes Place

In the afternoon of the second day instructors begin to integrate their knowledge of individual differences and the skills and processes involved in the helping relationship. This program is different because instructors are simultaneously learning problem identification and problem solving techniques. The concepts of personal learning and task orientation are combined.

Students with real school related problems become part of the third day. All students are given the Myers Briggs and the results are given to them beforehand. Students are informed of the nature of the process and human subjects regulations are followed. Students are all volunteers and are instructed that they are part of an instructor training program.

Instructors are given some background information on students as well as their MBTI scores. Three instructors work with one student; one person's role being the intervention process, the other two are observers. Following a 30-minute session students are dismissed and interviewed on a survey schedule that mirrors the skills in the 2½ day training program. Meanwhile the observers give the intervention instructor feedback. In a general debriefing session all groups are debriefed and student feedback is shared in a summarized form. The learning possibilities are enriched by sharing among instructor groups. The effects of instructor and student style, as measured by the Myers Briggs, are analyzed and discussed. Communications in this session are conducted in an impersonal way with the student's name and history protected.

Organizational Aspects

The school counselor(s) should experience the program before it is given to instructors. School administrators and curriculum specialists also should be participants. The opportunity to use concepts such as listening, feedback, and problem solving have organizational as well as instructional benefits. Because change is enhanced by a top-down approach, there are practically no good reasons for an attempt at training from the bottom up. Where training programs get little support in the permanent system from administrators and support personnel, few lasting results can be expected.

What makes this program successful is the research base, the field tested sequence, and the practical experience provided at the end of the program. A naval communications school in our West Florida service area has been using the program for two years. Some data exists supporting decreased attrition rates and extensive data exists in the form of st-training evaluation forms indicating that 95% of the participants rate the program as very successful.

A similar design has been utilized with chemical engineers with the goal being improved supervisor/employee relations. A study conducted in this environment indicates even higher participant ratings at subsequent times after training (i.e., post training, two weeks after, three weeks).

References

Joyce, B., & Weil, M. *Models of teaching.* Englecliff, N.J.: Prentice Hall, 1980.

Jung, C., Peno, R., & Emory, R. *Research utilizing problem solving (RUPS).* Portland: Northwest Regional Educational Laboratory, 1972.

Myers, I. B., & Briggs, K. C. *Myers Briggs type indicator.* Palo Alto, CA: Consulting Psychologists Press, Inc., 1962.

AD P000845

ARMORED FIGHTING VEHICLE IDENTIFICATION TRAINING:

A NEW PERSPECTIVE

by: Robert R. Begland Ph.D.
MAJOR, USA
CGSC
Ft Leavenworth, KS 66027

The complexity of the modern battlefield suggests that the chaos, confusion and destruction of previous wars will be only an inkling of the furor of tomorrow's potential conflict. It is incumbent upon the Army's leaders to quantify those factors that will contribute to success or guarantee failure on this future battlefield. The study described in this paper relates to one fundamental issue of fighting on the modern battlefield: the recognition and/or identification of armored fighting vehicles.

Historically, armies have produced weapon systems that have altered the ways in which wars are fought. One significant recent example of these changes is the engagement of armored vehicles. Due to the increased sophistication of weapons and range finders, various weapons are now capable of engaging and destroying armored vehicles at distances not previously possible. Yet the engagement and destruction of armored vehicles at extraordinary ranges is not practical unless the right vehicle is selected and engaged.

At the suggestion of the Director of Army Training, the Army School of Training Support of the Royal Army Educational Corps conducted a feasibility study into the issues associated with armored fighting vehicle recognition training. This feasibility study found that the present level of AFV Recognition/Identification Skills in the British Army were generally unsatisfactory and dysfunctional. Both of these findings were derived from a series of related Army-wide systemic deficiencies:

- o Recognition Training is generally accorded low priority in units;
- o AFV training is unrealistic;
- o AFV training focuses on a specific AFV with the salient features being taught, normally not observable at actual recognition distances;
- o the training media as utilized produced AFV recognition skills not readily transferrable to field conditions;
- o the training system was person dependent, in that it relied on the enthusiasm or dedication of the CO or an instuctor to create adequate AFV training, there was no systematic Army-wide support or direction.

The prototype AFV instructional system developed by ASTS was a practical approach resolving the above identified AFV problems, in light of the associated issues that influenced the content, delivery system, packaging and utilization in the field. The system is based upon:

- o Progression - AFV training begins at the recruit stage and continues in the soldier's unit through basic and advanced levels.

- o Menu Approach - The actual content and standards reflect the operational needs of each arm, unit or equipment employment. The soldier or commander selects from the total recognition training menu those vehicles/variants that are required for effective job performance.

- o Realism - Training would be linked to weapon and equipment deployment and use, and latest tactical doctrine.

System Characteristics. The prototype instructional system represented the collective integration of the most appropriate training media presently compatible with the budget constraints of 1981. The system was composed of a series of 45-minute AFV training packages presented as discreet instructional covered units. Each lesson covered four to five vehicles and included an assessment at the end. Each lesson had a series of slides for each vehicle with the salient features graphically embellished to facilitate learning. It also included a video-segment of each vehicle where the vehicle was rotated thru approximately 180 degrees with graphic embellisment of those features used for recognition, followed by video-segments of the vehicle moving cross-country under actual tactical conditions. In addition to the visual materials, the system also included one to one-hundred scale models for training, charts for wall posters, workbooks for soldier use, and a number of additional training devices for motivational purposes.

The factors considered in building the prototype trainging system were research based. The basis for this system was found in the application of a small number of relatively straight forward, but important training principles:

- o Teach the "Right Features"
  The most important decision that is made during the development of any AFV training system is the decision on which features are to be taught. The essence of AFV training is found in the set of features that are taught for each vehicle. Traditionally there have been two problems associated with the features that are taught: they were wrong the features and there were too many features. Research has repeatedly pointed out that the amount of information that an individual can mentally deal with on a given subject in a relatively short period of time is fairly limited. Therefore the number of features taught to a soldier should be within this approximate range. Secondly the features that are taught should possess practical relevance to either the recognition or identification of that vehicle and to the tactical environment. Too often the features that are being taught are visible only at unrealistically close distances. If the wrong features are being taught in a relatively large number, then the efficiency/effectiveness of that training is never in doubt.

- Teach the Soldiers "How to See"

  One of the major problems with the traditional approach to AFV training has been the reliance upon oral stimulation of the brain as opposed to a visual stimulation. Soldiers have traditionally been told a lot about the AFV, but little attention has been devoted to helping the soldier actually see the features that were being talked about. Research has shown that individuals of a lower ability level require assistance in focusing their attention to the significant aspect/feature being presented. The high quality, blown up image of a May Day Parade shot of a T-72 does not provide the appropriate visual stimuli to soldiers to be able to see the real image and the recognition/identification features such that he will remember them and be able to use them. The problem of teaching soldiers how to see involves three distinct, but related factors: direct the attention of the soldier to the discriminative stimulus to which he is to attend, while deemphasizing all other extraneous features and quickly get the soldier from a highly cued situation to a realistic situation with a minimal cueing. This was accomplished in the prototype lesson by the use of professionally produced training aids that were systematically developed to a precise design specification. Keep AFV training "visual". It is essential that the designer of the AFV training system keep in mind the real operational requirement for the soldier in either recognizing or identifying AFV. The task is predominantly a visual task and the training system should accommodate the visual need. All too often it seems that the slides being shown are in support of the words being said. The script or dialogue that is used in each lesson must be developed as an oral aid to the visual image. The visual image must convey the message to be learned, the words should support not suppress the learning experience.

- Facilitate "Transfer of AFV Identification Skills to Real World"

  The worth of any training solution is best measured through testing on the job. The prototype lessons facilitated the transferrence of skills by the use of video materials which were tied integrally to the slide materials. Through the rotation of the images of the vehicle on video, the soldier was given a "Gestaltist" or wholistic view of the vehicle, thereby helping him to bridge the visual gap from two dimensional slides to the three dimensional world of real tanks. Subsequent to these video rotations, the soldier was then brought to the realistic stage of seeing actual AFV moving under tactical conditions and distances.

- Help the soldier "remember what you taught him"

  The problem with teaching a subject such as AFV recognition is that the soldiers will forget it unless you do something to keep the proficiency level high. The prototype system demonstrated the various training opportunities that can be generated for soldiers to maintain the skills previously learned. The variety of revision activities include the individual soldier's AFV workbook that was designed to generate the mental rehearsal of recognition/identification features

during the critical period of five to eight hours after being instructed, AFV cards, slide rules, posters, and formal revision lessons. These devices and activities were an essential component to the system, in that the system must help the soldier remember what he has previously been taught.

o <u>Teach the unit instructor the "right way" to teach AFV Recognition/Identification skills</u>
Given the operational requirement for individual soldiers to be able to recognize/identify a relatively large number of AFV, it is obvious that the majority of AFV training is going to occur in the unit. Therefore, it is essential that the system be compatible with any problems and needs of units. One historical problem in the unit is that the instructor conducting AFV training will normally not be an accomplished AFV instructor. Regardless of the quality of the prototype material, the ultimate delivery of instruction will be achieved by the unit AFV instructor. It is necessary that the unit instructor be trained in the best way to teach, using the professionally produced AFV instructional materials.

In summary, the philosophy that underpined the development of this training system was that for the system to be maximally effective and efficient the designer should focus upon five basic rules: teach the right vehicles, teach only the most appropriate features, teach trainees how to see the features, help trainees to focus attention on each feature in order to improve the chances of remembering that feature, introduce realism which would allow the transfer of training to the real situation, and have the trained instructors conduct AFV training.

<u>Instructional System Prototype</u>. The instructional system prototype was composed of a series of lessons: two prototype lessons (APC and TANKS), each teaching five different vehicles, one enrichment lesson entitled "Comrades in Arms", two alternative review lessons, and a progress test. The lessons on APC's and TANKS comprised a slide tape and video based lesson while the enrichment lesson, the diagnostic test and the progress test were entirely on video. In addition to these basic lessons there were posters, playing cards, slide viewers, workbooks, and sand table models incorporated as aids to retention and motivation. The instructional paradigm for the primary instructional lessons was developed such that the soldier would be shown a series of approximately 14 slides which would progress from close up to more distant views of AFV and avoid the traditional side view. The use of graphic embellishment techniques would help the soldier to see the feature that was being taught and help him to remember the information. After the slides, the vehicle would be rotated through 180 degrees with the key features again being embellished. This segment was then followed by video clips of the actual vehicle at engagement distances. When all five vehicles had been presented in this way there would be a short assessment session to confirm that learning had in fact taken place.

<u>Design Rule for Slides</u>. From the outset the design of the slide presentation sequence was governed by eight basic rules. The title slide or first slide in the series for each vehicle should be a clear, close up view of

the vehicle with its name prominently displayed. This insured that the trainee quickly associated the vehicle with its name or identification number from the beginning. The second slide or scale slide in the sequence should present the immediate impression of relative size of the vehicle. This was achieved by the comparison of the vehicle to a man approximately six foot tall. The key features for each vehicle were then presented in their relative order of significance. Significance can be considered as a combination of prominence and permanence (e.g., can the feature be easily seen and is it unlikely to be removed or shot off?). Thus, the first feature that a soldier was taught was, in effect, the most significant discriminative stimlus available for him to either recognize or identify that vehicle. The fourth design rule relates to view. Side views of AFV are the most easy to identify and should be used sparingly. Front views are the most difficult and should be included. For the remainder of views a variety of frontal oblique views should be used. The availability of 35mm slides of real vehicles should not become the basis for instructional purposes. When teaching soldiers to see a particular feature, that feature should be presented clearly against an insignificant representation of the whole vehicle. This was accomplished using line drawings or photographs of scale models that were then graphically embellished. Concerning the graphic embellishment techniques, when teaching a specific identification feature, graphic techniques were used to highlight or embellish that particular feature so that the learner's attention was focused on it. The first slides in the sequence used this additional cueing strategy, but the cues were subsequently removed as the sequence progressed. The seventh design rule related to the review activities, where a review slide summarizing the features taught was presented after every three or four features had been taught and all the features were summarized on a slide towards the end of the sequence. The slides were designed to cause the soldier to mentally rehearse those features which he had just been taught and to associate those features with the name or number of that vehicle. The final design feature was the use of a tactical view, in that the last side should be a tactical view of the real vehicle in a real tactical study. These eight rules became the basis for the designing of the views that were used as the primary instructional strategy for teaching the initial identification skill.

Acknowledging the constraints of the paper and the size limitations, I will not address the strategies by which the salient features used for recognition and identification training were selected, other than to state that four independent strategies were developed and operationalized and that their combination produced the features that were taught for each AFV. Similarly, space does not permit a description of the actual techniques used for selecting the graphic embellishment procedures used in the instructional paradigm.

Evaluation Procedures. The AFV identification training system was subjected to both formative (developmental) evaluation and to summative (validation) evaluation. The developmental evaluation occurred throughout the prototype development effort, and every aspect/component of the system went through some form of developmental trials. The validation phase consisted of a large scale unit trial, in which the effectiveness of the total system was

demonstrated on representatives of the ultimate target population. In addition to this unit trial based objective data, extensive subjective data was collected from both the instructors who conducted the unit trials and the soldiers who participated. The test used to measure the performance of the recognition and identification skills of the soldiers was composed of 60 different test views (40 still photographs and 20 moving shots of actual AFV shown on a 22 inch TV.)

The Unit Trial Results. The performance test was administered to four Army units with total of 215 soldiers participating. The retention test was administered to three of these units with a total of 164 soldiers participating. The average score on the performance test for recognition was 88 percent correct and for identification 72 percent correct. The average score on the retention test for recognition was 92 percent correct and for identification 81 percent correct. The pretest scores were 46% recognition and 10.5% for identification.

Conclusions. This prototype training system achieved an operational demonstration of one approach to solving the poor standard of AFV recognition/identification skills in the British Army. The prototype resolved the problems identified in the earlier feasibility study (conducted by the author), while achieving quality AFV training in operational units, using the personnel assigned to those units. The prototype resulted in the production of a design specification of a workable AFV identification training system, that was compatible with the needs of the Army. It is important to realize that the results reported herewith were achieved as a result of the total system being used in the fashion in which it was intended. The various learning events were carefully planned and integrated to achieve a synergistic effect upon the soldier. If components of the system were to be removed and utilized in a strategy different than intended, the efficiency and/or effectiveness of that bastardized application must be doubted. The prototype training system represents an effective example of the application of systems theory to the identification and resolution of a legitimate performance problem facing the British Army.

The concurrent needs assessment effort which was also conducted as part of this total training system achieved the operational specification of the performance requirements of all soldiers in terms of their AFV recognition and identification training needs as perceived by their respective ARMS and services and as applied to their unique geographic operational position.

NOTE:
Individuals desiring more information concerning either the prototype instructional system or the needs assessment effort should contact either the author or the Commanding Officer, The Army School of Training Support, Royal Army Educational Corps Center, Wilton Park, Beaconsfield Bucks, England, HP92RP.

AD P008 46

OFFICER AS A COUNSELOR IN ELIMINATING SELF-DEFEATING BEHAVIOR

CONRAD G. BILLS, Captain, USAF, Assistant Professor of Aerospace Studies, Air Force ROTC, Washington State University, Pullman, WA 99164-2606.

JONATHAN M. CHAMBERLAIN, Ph.D., Counseling Psychologist, Counseling Center, Brigham Young University, Provo, UT 84602.

Much has been written on the officer's (commissioned or non-commissioned) role as a counselor. Zavachi and LaSota (11) noted "the Air Force cannot expect its present-day leaders to be counselors in the professional sense . . . However, it can expect today's commanders and supervisors at least to be familiar with certain behavior concepts and apply them in management of today's personnel force." Elsewhere the helping relationship (3) was discussed and the conditions for success in counseling were enumerated. Transactional analysis was offered as a tool for the effective manager to use in analyzing transactions with employees (9). Also, reality therapy or "RT" was presented as a practical approach in counseling subordinates (1). Yet in all of these discussions, no attempt was made to offer effective self-applied principles of behavior change.

The purpose of this article is to present effective principles for behavior change. These principles, when taught in the suggested sequence, can be applied by a person to change a self-defeating aspect of his/her life. The officer in the counselor role who knows these principles can use them in developing understanding of a person's (the helpee's) behavior as well as providing him/her (the helpee) a process for bringing about change. In effect, this paper presents several tools useful to the officer as a counselor. These tools are presented for the layman's application. Chamberlain (5) has applied these principles successfully in a home study program for Eliminating Self-Defeating Behaviors (ESDB). Self-defeating behavior (SDB) is defined as any recurring thought, feeling, or action that in some way prevents the doer from being a fully functioning person. There are many defeating behavior patterns ranging from deviant, aggressive sexual behavior and other forms of violence to feelings of timidity and shyness. These behaviors are exhibited in and out of the work environment and are often assessed as hindering work performance or hampering accomplishment of the mission.

The ESDB Program has been empirically researched and reported (2, 4, 6, 7, 10). Typical of non-empirical results is an immediate and one-month follow-up survey (Table 1, from twenty-nine school district personnel in a two-day ESDB workshop. They rated themselves on the following scale:

1. Wow! I no longer do my SDB!
2. Considerable change, but not completely eliminated
3. Noticeable change
4. Very little change (some but not much)
5. I do my SDB just the same as before the course (no change)

On the post survey, 81 percent reported a change (1-3 on scale above), and on the follow-up survey 90 percent reported a change (1-3 on scale above).

The same scale was administered at the conclusion of the workshop by 108 university students, mixing 9 separate groups conducted by 4 different leaders (Table 2). Ninety-two percent of the students reported a change with no significant difference between counselor groups, indicating the success was due to the principles presented and not the personality of the leader.

Again using the same scale, a summary of the self-ratings from 46 home study students completing the ESDB course to June 1976 (Table 3) indicated 94 percent of the students reported changes in behavior.

TABLE 1. Survey results from twenty-nine school district personnel.

Post April 19, 1975					
# Students (Scale)	1	2	3	4	5
26	7	7	7	5	0
PERCENTS:	27%	27%	27%	19%	0
	$\underbrace{\text{27\%}\quad\text{27\%}\quad\text{27\%}}_{\text{81\% changed}}$				

Follow-Up May 30, 1975					
# Students (Scale)	1	2	3	4	5
29	5	14	7	3	0
PERCENTS:	17%	49%	24%	10	0
	$\underbrace{\text{17\%}\quad\text{49\%}\quad\text{24\%}}_{\text{90\% changed}}$				

TABLE 2. Survey results from 108 university students.

Counselor-Leader	#	1	2	3	4	5
Counselor A	68	12	34	15	4	3
Percents	100%	18%	50%	22%	6%	4%
Counselor B	11	0	6	4	1	0
Percents	100%	0%	55%	36%	9%	0%
Counselor C	7	0	5	2	0	0
Percents	100%	0%	71%	29%	0%	0%
Counselor D	22	2	12	7	1	0
Percents	100%	9%	55%	32%	4%	0%
TOTAL	108	14	57	28	6	3
PERCENTS	100%	13%	53%	26%	5%	3%
		$\underbrace{\text{13\%}\quad\text{53\%}\quad\text{26\%}}_{\text{92\% changed}}$				

TABLE 3. Survey results from 46 home study students.

Number of Students	1	2	3	4	5
46	13	24	4	1	2
PERCENTS	33%	52%	9%	2%	4%
	$\underbrace{\text{33\%}\quad\text{52\%}\quad\text{9\%}}_{\text{94\% changed}}$				

The ESDB Program

Specifically, the ESDB program is a seven-step process designed to provide a person with techniques and personal assistance for use in eliminating behaviors that are self-defeating to them. The objectives of the program are three-fold: 1) to take a person through a step-by-step process designed to eliminate self-defeating behavior (SDB), 2) to help that person demonstrate control over that SDB, and 3) to help experience life without this SDB. Associated with these results is a discovery of the self as a person of worth and dignity and the casting aside of a worn, erroneous self-image. Some SDBs that have been eliminated by participants in the ESDB program are: inferiority feelings, compulsive eating, procrastination, fear of people, perfectionism, depression, alienation of others, and avoidance of reponsibility.

As a person enters the ESDB program, he/she is made aware that the same behavior identified to be eliminated is often used to defeat the change process. Some of these "defeating-the-program behaviors" are: being noncommittal to the change program, not fulfilling the assignments given, and putting the responsibility for change entirely on others. A key for the officer counselor is to obtain from the person a true commitment to change.

The seven steps in the ESDB program are: 1) How do I do my SDB? 2) How do I disown responsibility for doing my SDB? 3) What prices do I pay for doing my SDB? 4) What choices do I make to activate my SDB? 5) what negative techniques do I use to activate my SDB choices? 6) What fears must I face to be me without my SDB? and, 7) Facing my fears and discovering my inner self (5. 6).

A personal diary is kept by participants in conjunction with each step in the change program. The diary assignment is to record thoughts, feelings, and actions regarding the self-defeating behavior. Before going on to the next step the officer counselor critiques the diary by emphasizing personal responsibility for choices to do the behavior by interjecting in the submitted diary phrases such as: "I decide to" or "I told myself" or "I make myself feel" or "I choose to." A brief explanation of each step in the ESDB program follows. The optimum time frame to apply each step is 5 to 7 days.

Step One: How do I do my SDB? Here the person determines one SDB to eliminate. The choice should be the SDB that seems to be at the heart of problems one creates for oneself. Once the behavior is identified, then the daily diary assignment begins. From this diary the person can make a list of the ways he/she does this SDB. Then the individual can catch himself/herself doing, or about to do, the SDB and describe in the diary these thoughts, feelings, and actions.

Step Two: How do I disown responsibility for doing my SDB? By now the person is ready to identify and state the negative label, conditions, or "facts" about oneself, which he/she has been living under. These negative labels are warped self-images, images warped from the mirrors of life because of the SDB. These negative labels, although distorted and erroneous, are adopted as ones own because he/she does not know any better. Negative labels can be used as "cop outs" so a person feels he/she does not have to try. They feel they can blame the label, "I'm no good." Changing the negative label to a positive label can be done by shifting the "I am" personal connotation such as "I am a liar" to the "I do" behavior connotation: "I do lying." Another method is to state the opposite of the negative label such as "I eat sensibly." The new label can be a coined phrase or acronym and should be placed on the mirror at home, on the desk, in the lunch box, and other places where the positive label can be recalled and reinforced. Another method for disowning responsibility for the SDB is to shift the blame for doing the SDB to someone or something else. The person lists who or what is blamed in order to perpetuate the SDB.

Step Three: What prices do I pay for doing my SDB? Here the person makes an exhaustive list of the long- and short-range prices he/she pays for doing the SDB and the positive things missed. The importance of this step is to come to an understanding at a deep level the penalty paid for doing the SDB, convincing himself/herself that the SDB must be terminated. As a part of the diary assignment is a discovery and list of the methods used to minimize the price. The person sees the control he/she has over his/her own behavior. When the assignments are complete and the critiquing is done, the person can go on to step four.

Step Four: What choices do I make to activate my SDB? It is important to distinguish between two concepts of choice, i.e., the inner and the outer choices. The inner choice is a predetermined decision made prior to the time the person enters the situation where the SDB is usually done. On the other hand, the outer choice is the decision to carry out the inner choice. The outer choices are comprised of actions demanded by the situation. For example, if a person's inner choice is: "I will not trust my own judgment," then the outer choice is required: "I manipulate other people to make the decision on this issue for me." In step four the person recognizes and states the inner choices used to activate the SDB as well as the outer or action choices to help the SDB "happen" again. During this process the person can think through inner choices which allow him/her to pre-experience taking control over the SDB. The

person is shown a road map of life, and SDB route and a non-SDB route, so the person can see where he/she is at any given moment. Chamberlain (6) lists five major choices which a person makes to keep the SDB going. All of these are within their control: 1) choosing to do the SDB, 2) using outer choices to carry out inner decisions, 3) choosing to minimize prices paid for doing the SDB, 4) choosing to become irresponsible and disowning long enough to do it again, and 5) choosing to abandon ones best self each time he/she does it. The person is requested to list alternative choices that can be made in place of the choices leading to the doing of the SDB.

Step Five: What negative techniques do I use to activate my SDB choices? The person having experienced the reality of being the chooser and doer of the SDB is likely to have discovered ways of keeping the SDB active. These subtle "aids" are called negative techniques. Some examples are comparing oneself to others, anticipating certain things might occur, distorting feedback, intellectualizing, pouting, manipulating oneself and others, blanking the mind so the problem cannot be dealt with realistically, and placing unreasonable expectations on oneself and others. These negative techniques are like fuel to a fire. They keep the fire burning; without them the fire would go out. So without techniques to keep it going, the SDB would cease to exist.

In step five the person lists 1) his/her negative techniques used to activate the SDB and 2) some positive techniques needed to be developed and used to keep on the non-SDB route. The person should also stop using the negative techniques, keeping a daily diary for critique on the struggles to change.

Step Six: What fears must I face to be me without by SDB? As a person grows from infancy, he/she responds to the world as an "integrated self." As new anxiety producing situations arise, the person chooses either to respond as the fully integrated person or to abandon this non-SDB route for methods which result in SDB patterns (6). These fear and anxiety feelings are perceived to such a degree that the person forsakes the integrated self. The person feels there is no other alternative in order to cope with the situation. Because these choices were made under stress, the behavior is deeply seated and based on an erroneous assumption. Thus, in present life, behavior is partially based upon fear, a fear of re-experiencing the original situation, but now stored away. This is the SDB creation story. However, the reason these learned SDB patterns continue is the fear of living without them.

In step six the person lists what he/she fears when considering letting go of the SDB. These fears can be grouped under what the person might find out about himself/herself or what might happen to him/her. Next comes the recognition that these fears are mythical in nature, contrived by the person to keep the SDB because they are least likely to occur on the non-SDB routes. Though these fears are distorted and erroneous, they are often perceived by the individual as awesome and foreboding. Finally, in this step is a counter effort, listing the positive benefits to be received by dropping the SDB.

Step Seven: Facing my fears and discovering my inner self. At this point, it might be helpful for the person to complete this step in the presence of a professional helper, a close friend, or a loved one who is willing to read aloud the "guided imagery" session (6). During this session through a simple process, the person faces the barrier to being ones best self. The officer counselor guides the person in a step-by-step manner through a mythical or imaginary barrier by having them close their eyes and imagine certain ideas or experiences. This is best accomplished in a location without interruptions for about thirty

minutes. Once the person has gone through the exercise, listening carefully and responding mentally and verbally, he/she writes these experiences and any insights learned. The typical person will face at deeper levels a mental barrier which prevented progress along the non-SDB route and do away with the barrier; identify an origin of the SDB and feel it is no longer needed. Also, the person will discard the negative self image and visualize coping in life without the SDB.

Beyond this step is an opportunity for the person to gain the realization he/she is not alone in the things imagined and receive additional help in understanding the relationship between the things imagined and the real-life struggles. Chamberlain (6) described this step as follows: "Breaking through the barrier seems to represent a symbolic breaking out of a long-held habit into a new life."

Adaptation for the Officer as a Counselor

The evidence is growing that the officer as a counselor can apply the Eliminating Self-Defeating Behavior (ESDB) program as an effective method for eliminating undesirable behavior. This method has been successfully applied in group and individual counseling as well as by home study (6). Chamberlain's cassette recordings (7) are useful to the lay counselor and the counselee. Since a person can go through this program almost on their own, the ESDB program could be integrated as a tool for the officer in the role of counselor.

The initial session is to identify the self-defeating behavior (SDB) to be eliminated and begin the ESDB program or refer the counselee to the home study course (5). Progress dates to complete the program steps are established. The officer counselor creates the atmosphere for a helping relationship in the ESDB program. The assignments for homework will include a daily diary. Next, the officer counselor sets an appointment with the counselee to evaluate the first step of the ESDB program. Interim follow-up might be needed depending on the counselee.

During the second session the officer counselor critiques the counselee's diary, discusses with the counselee how he/she disowns responsibility for doing the SDB, and identifies a negative as well as a replacement positive self-label. At this point the officer counselor assesses the progress and determines how many steps the counselee could accomplish on his/her own before the next session. Regardless of the number of steps accomplished between sessions, the officer counselor establishes regular intervals for the counselee to turn in homework. The officer counselor reviews the homework, makes helping comments, and returns the critiqued homework to the counselee.

Step seven of the ESDB program often requires additional attention on the part of the officer counselor. Here the cassette tape recording or the step-by-step procedure in the ESDB text can be followed. Perhaps the officer counselor would feel a need for referral to a counseling agency at this stage or the change process.

It is important to [illegible]ize that the ESDB program is an effective tool. It offers a catalyst for change. The program teaches a method for understanding how a behavior is self-defeating, the choices that can be made to eliminate it, and the mythical barriers that can be discarded so the integrated self can operate.

Some results of applying the ESDB program have been demonstrated. The change in the counselee's life brings about a chain of events that improves his/her performance in and out of the work environment. The counselee's self-esteem is enhanced. The overall result is the preservation of our most valuable resource, our people. The time investment for this process is minimal compared to other alternatives of dealing with personal behavior problems. Though time requirements vary between counselees, the key point is the self-help nature of the ESDB program. Some counselees have successfully eliminated their SDB without any assistance in formal counseling sessions.

The ESDB program is a tested counseling procedure. It has direct application in the role of the officer as a counselor. The published materials make the program useful to the lay counselor. The steps are sequential, the homework assignments are established, and the methodology for feedback is well outlined. The cassette tapes and the home study course even offer a professional counselor to assist in the process of eliminating self-defeating behavior when needed.

References

1. Air Command and Staff College. Describing reality therapy as a practical approach in counseling subordinates. Course 30B, Lesson 15, Extension Course Institute, Air University, Gunter AFS AL, 1981, 11-19.

2. Banks, J., Grimmer, J., Hardy, R.E., Hiatt, D., and Lowe, J. Self-defeating behavior workshops: Systems approach for hard-to-serve veterans. Personnel and Guidance Journal, 1979, 57, 313-315.

3. Benjamin, A. The Helping Interview. Boston, Mass.: Houghton Mifflin, Co., 1969.

4. Cudney, M. Eliminating Self-Defeating Behaviors. Kalamazoo, Mich.: Life Giving Enterprises, 1975.

5. Chamberlain, J.M. Eliminating a Self-Defeating Behavior: A Brigham Young University Home Study Course (Ed. 514X-1, 4th ed.). Provo, Utah: Brigham Young University Press, 1974, 1976.

6. __________. Eliminate Your SDBs (Self-Defeating Behaviors). Provo, Utah: Brigham Young University Press, 1978.

7. __________. Defeat Your Self-Defeating Behaviors. Salt Lake City, Utah: Covenant Recordings, 1980.

8. Johnson, E.K., and Chamberlain, J.M. The treatment of smoking as a self-defeating behavior. J. of Psychology, 1978, 98, 37-43.

9. Luchsinger, V.P., and Luchsinger, L.L. Transactional analysis for managers, or how to be more ok with ok organizations. MSU Business Topics, Spring 1974, 5-12.

10. Parks, C., Becker, W.M., Chamberlain, J.M., and Crandell, J.M. Eliminating self-defeating behaviors and change in locus of control. J. of Psychology, 1975, 91, 115-120.

11. Zawacki, R.A., and LaSota, P.E. The Air Force supervisor: Giving and receiving help. Air University Review, 1974, 25, 78-82.

Combining Results of Independent Research in Tank Crewman Performance

Barbara A. Black
U. S. Army Research Institute

Charlotte H. Campbell
Human Resources Research Organization

The effectiveness of any combat weapon system is in large measure a function of the level of performance of the soldier operator. The US Army is interested in ensuring maximum effectiveness of the new M1 tank system by optimally selecting and training M1 crewmen. To support the Army's effort in maximizing system effectiveness from the personnel selection aspect, the Army Research Institute has conducted extensive research in the area of tank crewman performance prediction during the past several years. The purpose of this paper is to evaluate the results of this research in order to determine by job content area, for both trainees and job incumbents, whether quantifiable aptitudes are related to tank gunnery performance.

Black and Kraemer (1981) identified three aptitude categories which potentially underlie gunnery performance. These included a cognitive component as encountered in troubleshooting, a perceptual component as in target acquisition and a psychomotor/perceptual-motor component as in target tracking. Each of the four crew positions within the tank system (i.e., loader, driver, gunner and tank commander) requires performance of tasks which appear to contain these components, albeit in varying degrees. A review of the Armor crewman performance prediction literature lends support to this categorization but points to an additional dichotomy with reference to research techniques utilized. Techniques include paper-and-pencil tests as well as tests called job samples which require either simulators or actual tank equipment.

These aptitude categories and research techniques were identified in the tank crewman performance prediction literature. In the area of cognitive testing, the literature included validation of ASVAB-derived composite scores such as CO, GT and AFQT as paper-and-pencil predictors of gunnery performance (Greenstein & Hughes, 1977; Campbell & Black, 1982; Black, in preparation), and simulator based tests of the tank fire control computer (Campbell & Black, 1982; Black, in preparation). For perceptual testing, paper-and-pencil tests are also the most commonly encountered (Greenstein & Hughes, 1977; Eaton, 1978; Eaton, Bessemer, & Kristiansen, 1979), although two instances of simulator based perceptual tests were found (Eaton, Johnson, & Black, 1980; Campbell & Black, 1982). Validation of psychomotor tests using hands-on equipment can be found in three reports (Eaton, 1978; Kress, 1980; Black, in preparation), and finally, simulation techniques are applied to psychomotor performance prediction in two reports (Eaton et al., 1980; Campbell & Black, 1982; Black, in preparation). The correlations reported for these research efforts provided the data for the meta-analyses.

Method

The eight documents included in the review of Armor crewman performance prediction literature produced a total of 18 data sets for evaluation. Data sets were accepted for meta-analysis based upon the following criteria:

1) predictor variables were obtained from tests which could be classified as either cognitive, perceptual or psychomotor/perceptual-motor, 2) criterion measures were tank live fire gunnery hit scores, and 3) subjects were either tank gunner trainees or operational unit gunner/TCr.

Data sets were placed into analytic categories according to the format presented in Table 1. Each data set had between one and ten correlations that were used in the meta-analyses for each analytic category.

Table 1
Number of Correlations (and Data Sets)
Available for Meta-Analysis

APTITUDE CATEGORIES	TEST TYPE Paper-and-Pencil	Job Sample
Cognitive	18 (11)	8 (2)
Perceptual	63 (10)	6 (6)
Psychomotor or Perceptual-Motor	-	41 (11)

Two methods were used for combining and evaluating the results reported in the literature. The first, drawn from Rosenthal (1978), used exact probabilities (one-tailed) of the correlations to compute an overall Z for each data set; the exact probabilities were corrected for the number of correlations drawn from each data set in each analytic category. The Z-values for each data set in each category were then combined using a method whereby each Z is weighted by the degrees of freedom of its respective data set. The method yields a Z for each analytic category (see Table 2).

The second method was based on Glass (1977), who advocates the averaging of correlation coefficients or coefficients of determination. Here, the Fisher z-scores were computed for each correlation and combined (Snedecor & Cochran, 1967) first within data sets and then across data sets within each category to yield an overall weighted average z. This value was then converted back to a correlation; the squared correlations, representing the proportion of variance accounted for, are reported in Table 2.

Results and Discussion

While the aggregated results of cognitive paper-and-pencil testing produced a statistically significant cumulative Z for trainees, it is interesting to note that the average variance in gunnery scores accounted for by the cognitive component is only 2.5%. So although the predictions are consistent and reliable, they do not provide very much information. One variable of the cognitive job sample tests, computer accuracy, was a significant predictor for operational unit personnel, accounting for over 10% of the variance in gunnery performance, but the variable is not a significant predictor for trainees.

Table 2
Results of Two Meta-Analysis Techniques Relating
Tank Crewman Aptitudes to Tank Gunnery Performance

	OPERATIONAL UNIT SOLDIERS		TRAINEES	
	Z	Variance Accounted for	Z	Variance Accounted for
COGNITIVE				
Paper & Pencil Tests	1.511	2.5%	2.171*	2.5%
Job Sample Tests				
.M1 Computer Accuracy	2.106*	10.6%	.977	0.4%
.M1 Computer Speed	-.627	2.5%	1.079	0.8%
PERCEPTUAL				
Paper & Pencil Tests	-6.741	0.0%	-4.957	0.1%
Job Sample Tests				
.Round Sensing	-	-	3.002***	4.8%
PSYCHOMOTOR/PERCEPTUAL-MOTOR				
Job Sample Tests				
.Tracking Accuracy	1.441	6.0%	1.266	1.0%
.Tracking Speed	-.122	0.0%	1.388	0.9%
.Main Gun Lay Accuracy	2.542**	7.1%	-	-
.Main Gun Lay Speed	2.239*	4.9%	-	-
.Target Engag. Hits	-.433	0.1%	.581	0.1%
.Target Engag. Speed	-	-	.245	0.0%
.Sub-Caliber Hits	-.239	0.1%	-	-
.Sub-Caliber Speed	1.547	4.0%	-	-

*p < .05 one-tailed
**p < .01 one-tailed
***p < .001 one-tailed

Perceptual paper-and-pencil tests were poor predictors of gunnery scores for both operational unit personnel and trainees. The job sample test approach, however, produced positive correlations in all data sets, for a highly significant effect, but the variance accounted for averages less than 5%. Whether the approach would be effective among operational unit personnel is unknown. Two of the job sample tests of psychomotor/perceptual-motor aptitude were significant predictors across studies for operational unit soldiers, but none was a predictor for trainees.

Overall, it would appear that job sample tests are better predictors of performance by job incumbents than are paper-and-pencil techniques. For trainees, however, where performance is usually measured during their earliest experience on the tank, hands-on tests are sometimes predictive, and so are paper-and-pencil tests. It should be noted that no attempt was made to separate concurrent predictions and actual time-separated predictions for the analyses of unit personnel performance. And because perceptual paper-and-pencil tests were combined within data sets and adjusted for that process of combining, the large numbers of small correlations in each data set caused the combined Zs for the sets to be very large negative numbers. Examination

of individual tests across research efforts could lead to different conclusions for a few. In general, meta-analysis techniques appear to be valuable tools in assimilating independent research results and providing insight for future research efforts.

References

Black, B. A. M1 Crewman Gunnery and Training Performance Prediction (Research Product). Fort Knox, Kentucky: US Army Research Institute for the Behavioral and Social Science, in preparation.

Black, B. A., & Kraemer, R. E. XM1 Gunnery Training and Aptitude Requirements Analyses (Research Product 81-5). Fort Knox, Kentucky: US Army Research Institute for the Behavioral and Social Sciences, 1981.

Campbell, C. H., & Black, B. A. Predicting Trainability of M1 Crewmen (Final Report FR-MTRD(KY)-82-7). Fort Knox, Kentucky: Human Resources Research Organization, 1982.

Eaton, N. K. Predicting Tank Gunnery Performance (Research Memorandum 78-6). Fort Knox, Kentucky: US Army Research Institute for the Behavioral and Social Sciences, 1978.

Eaton, N. K., Bessemer, D. W., & Kristiansen, D. M. Tank Crew Position Assignment (Technical Report No. 391). Fort Knox, Kentucky: US Army Research Institute for the Behavioral and Social Sciences, 1979.

Eaton, N. K., Johnson, J., & Black, B. A. Job Samples as Tank Gunnery Performance Predictors (Technical Report No. 473). Fort Knox, Kentucky: US Army Research Institute for the Behavioral and Social Sciences, 1980.

Glass, G. V. Integrating findings: The meta-analysis of research. Review of Research in Education, 1977, 5, 351-379.

Greenstein, R. B., & Hughes, R. G. The Development of Discriminators for Predicting Success in Armor Crew Positions (Research Memorandum 77-27). Fort Knox, Kentucky: US Army Research Institute for the Behavioral and Social Sciences, 1977.

Kress, G. Validation of Tank Gunnery Training Tasks (Draft Final Report MDA 903-78-C-2042). Alexandria, Virginia: Human Resources Research Organization, 1980.

Snedecor, G. W., & Cochran, W. G. Statistical Methods (6th ed.) Ames, Iowa: Iowa State University Press, 1967.

TESTING A HALF MILLION TRAINEES: AFMET PROJECT

Wallace Bloom Ph.D
Wilford Hall USAF MEDICAL CENTER (SGHMSR)

To facilitate early identification of basic trainees with significant psychiatric problems, the Air Force Medical Evaluation Test (AFMET) Project has been conducted by the Wilford Hall USAF Medical Center since June 1975. (Chart 1) (Bloom 1980b). By 30 September 1982, the initial screening (Phase I) had processed 532,384 basic trainees at Lackland Air Force Base. Six and a half (6½) percent were not cleared for return to duty (RTD), and these 34,462 were identified for individual interviews and further tests (Phase II). After this, 8,055 were still not cleared for return to duty and more than a third of these were recommended for discharge after being clinically interviewed, tested, and diagnosed at Phase III (Table 1). In fiscal year 1982, the USAF saved $6,340,430 by release of those trainees who were neither suitable for nor adaptable to USAF duties and military life. Since the project has been described in other papers and presentations, this article focuses on the evolutionary dynamic changes during the first seven years to accomplish the fine tuning suggested by General B. Davis. As data was collected and analyzed, improvements became possible and these were implemented as soon as feasible.

Phase I: Initial Screening by a Computer Scored Test.

Initially (1975) all incoming basic trainees where given the Historical Orientation Inventroy (HOI) at the Lackland AFB Reception Center within an hour of their arrivals (Bloom 1977a). This often occurred late at night and the procedure was changed in October 1976, to testing during normal duty hours on the second day of training (2-DOT) in classrooms adjacent to the eleven squadrons. The original 100 item HOI test (Guinn) was changed to 50 items to reduce errors when marking responses on optical scoring sheets and eliminate fifty of the unscored (camouflage) items. In 1982, the items were printed right on the response sheet rather than on a separate card, and some demographic data was added. A study of the over forty thousand 1977 enlistees, tracked for 4 years to identify factors related to early attrition due to unsuitability, and/or unsatisfactory behavior or performance, indicated (after stepwise multiple linear regression and other statistical analyses) that 28 factors could better identify trainees unlikely to complete enlistments. Each HOI question could be given an individual preductive weight rather than just one of two numerical scores. The items added as useful predicitive factors were education, age, sex, and marital status. These improvements were accomplished by 1 February 1982, along with change in data processing on the Sentry 60 system rather than OPSCAN 17 so that data went directly to magnetic tapes rather than to two IBM punched cards per individual. Members of the Air National Guard and Reserves were also identified as such. Data processing was shifted from the Human Resources Laboratory to Air Training Command resources as the Air Staff had directed the experimental project be made operational after they had reviewed the first year's data.

Phase II: Individual Interviews and Tests.

During the research year, all Phase II interviews and tests were at the Base Dispensary starting after supper, at 5:00 PM. In the fall of 1976, three satellite mental health clinics within or in close proximity to the Basic Military Training Squadrons were established as OUTREACH facilities, and the AFMET interviewers were integrated with other mental health personnel. This facilitated communications, and cooperation with training instructors and commanders. Phase II interviews, and tests were conducted at these clinics during normal duty hours (O'Hearn 1978).

Data from the interviews of each selected trainee (6½ to 7% of all trainees) were marked on a 5 point scale for 39 items by specially trained enlisted mental health technicians. Subsequent analyses produced periodic frequency distribution counts, inter-item correlations, differences in average scores of groups retrained or discharged, and correlation with psychological test scores (Bloom 1981).

The Minnesota Multiphasic Personality Inventory (MMPI) was first used as part of Phase II screening and the computer scoring took two days. It did not help very much and the clinicians in Phase III reported excessive misidentifications. Air Force norms were developed but their use resulted in little improvement. It was supplemented with a sentence completion test (Bloom Sentence Completion Survey, BSCS) which was taken by Phase II selectees prior to their interviews. The data proved to be useful as interviewers found responses furnished useful icebreaker information and helpful information about the subjects attitudes. Subsequently they learned how to score the seven subtests (People, Physical Self, Family, Psychological Self, Self-Directness Work, and Accomplishment) (Bloom 1975). The scores had high inter-rater reliability, stability over time, and validity as those returned to duty averaged 14 points higher than subjects recommended for discharge. Use of MMPI was moved to Phase III (Bloom 1980a).

Phase III: Clinical Interviews and Tests.

Usually between the 8th to 12th day of training, airmen not cleared for return to duty at Phase II are scheduled for clinical and diagnostic interviews by officers. During the past seven years, the use of psychological tests has increased and data optical scanning sheets have been redesigned so that test scores become part of the data bank. Clinicians select the tests for each individual and those used included: MMPI, Gordon Personal Profile, Firo B, 16 PF, TAT, WAIS, Shipley and Neuropsychologicals. The diagnoses most often used when recommending discharges were: (1) Atypical, mixed or other personality disorder, which includes immature personality disorder (301.89). (2) Adjustment disorder with mixed emotional features (309.28). (3) Avoidant personality disorder (301.82). (4) Dependent personality disorder (301.60).

Some trainees were found to have difficulty in coping with the stresses of training and needed brief supportive therapy. Therapy groups now are scheduled twice a week, and 91 trainees participated in October 1982. Attendees often had high state but low trait anxiety, and almost all were enabled to complete basic training.

Conclusions: Results and Benefits.

In 1 October 1982, a new Air Force Regulation 39-10 became effective which emphasized the substantial USAF investment in airmen, and that those "who do not show a potential for further service should be discharged" (Par 5-1-a). A condition which maybe a basis for discharge is a Personality Disorder supported by a report of evaluation by a psychiatrist, or a psychologist (Par 5-12-i). Since AFMET project focuses on trainees psychological problems that existed prior to enlistment, and contribute to poor adaptation to military life, this project is strengthened by the new regulation. It has saved the Air Force millions of dollars each year by early identification, and elimination of individuals who otherwise would have used up many more training dollars, pay, supplies, administrative costs before later being discharged prior to expiration of term of service (PETS), and rarely ever being productive. Unsuitable or unadaptable individuals have been spared the stresses, and emotional damages that might have resulted from their further retention. We have developed procedures for more effective use of enlisted mental health technicians.

Useful norms for standardized psychological tests have been developed. A computerized mental health data bank has been established for data collection, retrieval, analyses and reporting. This buildup of information, and normative data may, in future years, prove to be a major benefit of the project as we can learn the demographic, interview and test data variable that relate to success or failure. Not to be overlooked is the fact that more than half a million trainees have gone through the AFMET mental health screenings and none had committed suicide here in spite of the stresses of training and uprooting from home. A similar age population group in civilian life would have been expected statistically to have had at least seventeen suicides in the past seven years.

REFERENCES

Bloom, W. Manual, Bloom Sentence Completion Surveys; Student and Adult Effectiveness Training Associates of San Antonio, 133 Twinleaf Lane, San Antonio, Texas 78213, 1975

Bloom, W. Air Force Medical Evaluation Test, USAF Medical Service Digest 28, (2) 1977a.

Bloom, W. Relevant MMPI Norms for Young Adult Air Force Trainees. Journal of Personality Assessment 1977b 41 (5) 505-510.

Bloom, W. Manual: Bloom Sentence Completion Survey: Adult, Chicago, Illinois Stoelting Co. 1980a.

Bloom, W. Air Force Medical Evaluation Test (AFMET) Identifies Psychological Problems Early. USAF Medical Service Digest Nov-Dec 1980b, 8-9.

Bloom, W. Assessment of Patients: The Standardized Report of Interview (SRI). Military Medicine, 1981, 146(9) 639-641.

Bloom, W. The AFMET Project and the Bloom Sentence Completion Surveys presented at the Eighth Psychology in the DOD Symposium, USAF Academy, Colorado, 22 Apr 1982.

Guinn, N., Johnson, A. L. & Kanton, J. E. Screening for adaptability to military service. Report No. AFHRL-TR-75-30 Air Force Resources Laboratory, May 1975.

O'Hearn, T. P. & Bloom, W. The Air Force Medical Evaluation Test (AFMET) Program: A Three-Phase Assessment Program. USAF Behavioral Sciences Symposium, May 24, 1978.

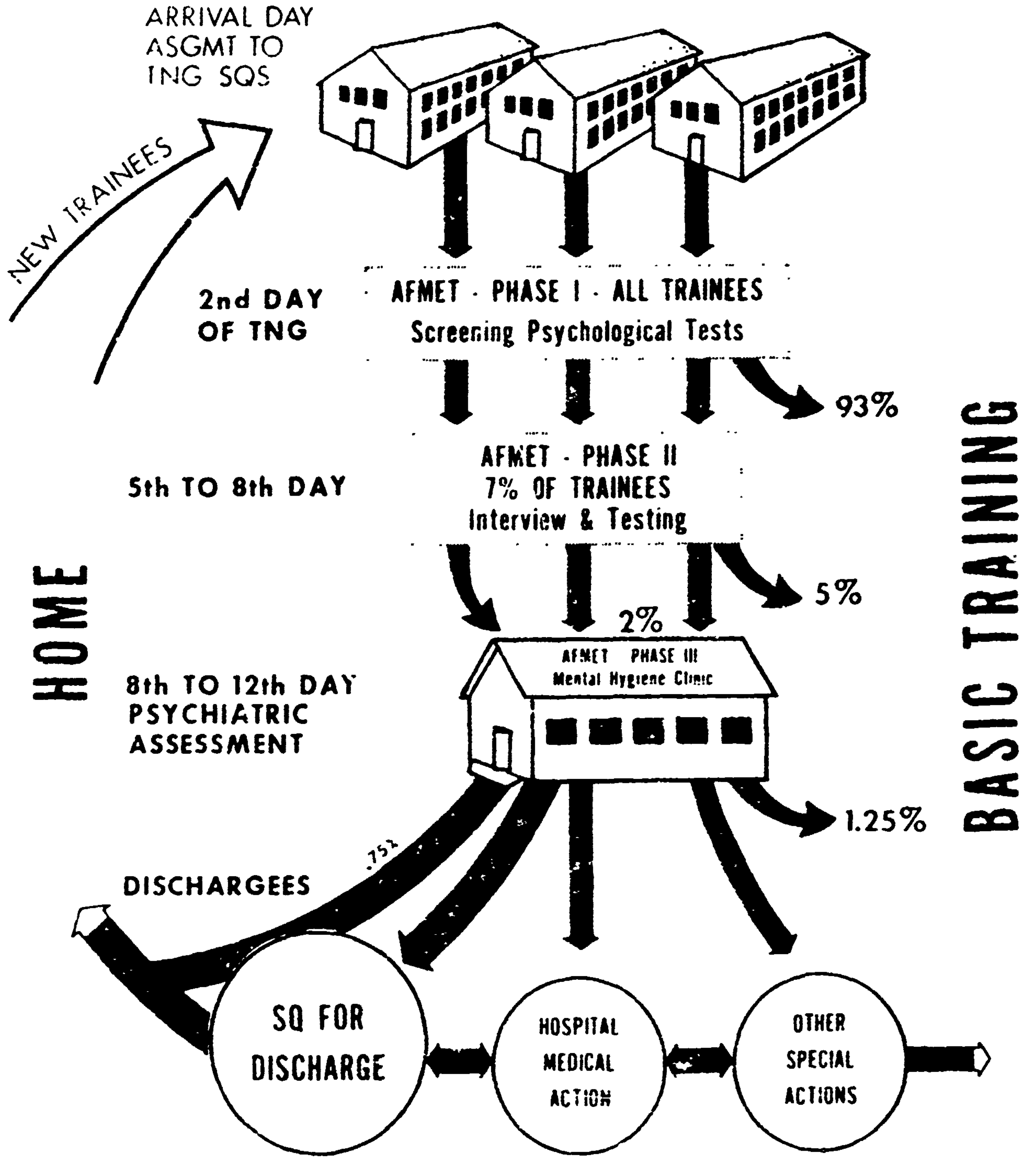


Figure 1. Diagram shows how AFMET screens new trainees through three-stage testing, sending a few back home (to left) if they would not be able to adapt emotionally to military life. After AFMET screening, more than 99% of new trainees enter basic training.

A F M E T Project

Mental Health Screening of USAF Basic Trainees

First Seven Years

	FY76 1 Jun 75 to 15 Jun 76	FY77 28 Sep 76 to 30 Sep 77	FY78 1 Oct 77 to 30 Sep 78	FY79 1 Oct 78 to 30 Sep 79	FY80 1 Oct 79 to 30 Sep 80	FY81 1 Oct 80 to 30 Sep 81	FY82 1 Oct 81 to 30 Sep 82	Total 7 Years
PHASE I, SCREENING								
HOI Test	80,732	73,666	72,235	70,227	77,487	83,120	74,917	532,384
Selectees	5,369	4,918	4,585	4,692	5,580	5,314	4,004	34,462
Percent Selected	6.7%	6.7%	6.4%	6.7%	7.2%	6.4%	5.3%	6.5%
PHASE II, Interviews								
RTD'S	3,941	2,725	2,773	2,986	3,820	3,585	2,598	22,428
To MHC Phase III	1,331	1,054	1,015	988	1,209	1,321	1,137	8,055
Prior Discharges	(NA)	(NA)	522	430	418	387	359	2,116
Not Accounted for Shipped Before AFMET Completion							91*	91*
Discharges of AFMET								
Selectees								
Recommended by PHASE III	306	512	303	251	369	380	421	2,542
Other (Sq, EPTS, etc)	134	325	571	538	531	535	359	2,993
Total Discharges of Selectees	440	837	874	789	900	915	780	5,535

*Since 1 Apr 82, new data processing procedures and forms, made these identifications feasible.

Maintenance Performance

Douglas J. Bobko

US Army Research Institute

During the next twenty years, there will be within the Army an unprecedented increase in both the number and sophistication of military systems. It is now important to recognize that modern and more efficient techniques are needed to manage the development and support of those Army personnel who operate and maintain these systems. The unique problems encountered in the development and support of personnel are illustrated below by comparing the life-cycle of military personnel with the life-cycle of military hardware. Following this comparison, a field-tested prototype management system (the Maintenance Performance System) which addresses these problems is discussed.

The life-cycle of a weapon system begins with a current or future mission requirement. Once the mission requirements are adequately specified, the weapon is designed, developed and tested. If the weapon does not perform to specifications, appropriate design changes are made and the system is reevaluated. When a satisfactory prototype is achieved, the system is mass-produced and fielded. To aid in the support of the fielded weapon system are performance measures such as probability of part failure, mean time between failure, mean down time, mean repair time, etc. Although these measures may be somewhat less than reliable, they establish an important standard against which support requirements can be anticipated. The Army-wide management of materiel is further aided by relatively rigorous and standardized data collection and reporting systems such as the Maintenance Control System (MCS).

The scenario above differs significantly from the development and support of personnel. Although personnel requirements are driven by and can be estimated from materiel characteristics, personnel cannot be mass-produced to meet those requirements. Rather, personnel developers (i.e., military trainers and educators) start with a heterogeneous group of recruits who differ in education, experience and motivation; probably none of these recruits arrive with any of the skills needed to operate or maintain military hardware. About the best that can be done at this point is classification of personnel according to more or less valid measures of aptitude, placement into Military Occupational Specialties (MOS) according to aptitude measures and manning requirements and, finally, enrollment for several weeks in an MOS-specific Advanced Individual Training (AIT) curriculum.

When "fielded," however, these soldiers are far from possessing more than the basic skill requirements for operating and maintaining hardware systems. Moreover, the development of any soldier's skill is never complete since the skill requirements change as a function of advances in grade, changes in equipment design, the fielding of new equipment, and unit missions.

Unlike military hardware, then, the development and support of personnel begins in large part AFTER they are placed into the field and continues for many years. According to Army doctrine, this development is to take place at the lowest organizational level and is to occur primarily in the form of supervised on-the-job training (OJT). However, unit level training managers and training supervisors have been provided with little in the way of guidance or resources to accomplish this training. The result has been a documented lack of proper and sufficient technical skill training. What is lacking at this point in the personnel life-cycle is an effective system to monitor the performance and utilization of personnel much as MCS monitors these factors for materiel.

As a first step toward addressing this problem in the management and delivery of unit level OJT, the US Army Research Institute (ARI) developed a prototype performance and training management information system for use at the direct support maintenance level (a more extensive system for use at the organizational maintenance level is under development). This system is called the Maintenance Performance System (MPS) and it is designed to identify training strengths and deficiencies, to locate available training resources, and to monitor the effect of training on job performance.

MPS is an automated maintenance management information system which provides to training supervisors up-to-date and unique information about WHO needs to be trained, WHAT tasks need to be trained, and HOW training can be accomplished. This information is presented in report form so that training opportunities can be easily recognized and taken advantage of within the context of a unit's available resources and constraints. Of equal importance is that MPS provides quantitative measures of individual and unit-level proficiency and efficiency (e.g., job completion time) so that the effects of training can be assessed.

Information for MPS is collected through the use of two simple input forms which are completed by technical MOS supervisors. One of the forms is attached to the job order packet and is used to record job performance data and OJT experience. The other form is used to record special training or the occurence of performance-based tests such as the Skill Qualification Test (SQT). Based on observations to date, supervisors spend about ten minutes each week completing these forms. A microcomputer is used to process the information and to print management reports.

As a means to improve the conduct and quality of unit level training, MPS is successful in several ways:

* ACCEPTANCE BY USERS: MPS has been operational for more than a year at two divisional FORSCOM maintenance battalions and is accepted by users as a system which provides timely, accurate and useful training-needs information.

* GUIDANCE OF TRAINING: MPS information is used to guide the course of training, to make job assignments, and to serve as a memory refresher about which repairmen require special training on critical skills.

* AUTOMATED JOB BOOK: MPS frees the training supervisor from making daily entries into each soldier's job book. As part of its regular report output, MPS provides each supervisor and individual repairman with an up-to-date record of OJT and other types of training.

* SKILL AND PERFORMANCE BANK: The historical records of skill and performance data provided by MPS constitute a skill bank from which battalion and company level management assess current unit proficiency and readiness.

* INCORPORATION INTO SAMS: MPS-like training information has been approved for incorporation into the Standard Army Maintenance System (SAMS). This is a milestone in that it represents the first systematic Army-wide collection of training and performance information, a vital step toward an improved and integrated training management system.

It should be noted that additional benefits accrue from the MPS data-base itself and that these benefits extend beyond the unit level. The maintenance performance information found in MPS can be used, for example, to target Army-wide skill deficiencies and fine-tune institutional training curricula, to pinpoint areas in which training materials need to be developed or improved, to estimate future manning requirements, to establish more reliable and comprehensive performance standards, to aid in the design of hardware, and to evaluate differences in training strategies and training management.

It is clear that the development of military personnel is different from that of hardware in two important respects: (1) skill development occurs largely after personnel are assigned to a unit, and (2) skill development continues for the duration of a military career. MPS was designed with these differences in mind and has been demonstrated to be an effective tool for the management of training and skill development. As we enter a period of declining available manpower and increased weapon sophistication, more attention will be focused on the quality of personnel and the need for systems such as MPS will grow.

VALIDATION STUDIES OF THE BELGIAN ARMED FORCES RESERVE OFFICER SELECTION SYSTEM

A. BÖHRER & H. LUYTEN
BELGIAN ARMED FORCES • PSYCHOLOGICAL RESEARCH SECTION • BRUSSELS

The actual Belgian reserve officer selection system proves to be valid. M[illegible] regression analysis of the predictor variables on the officer [illegible] outcome yields a substantial increase in prediction capacity [illegible] luced number of predictors : 'Intelligence' and 'Sense of Res-[illegible]lity' emerge as the most important ones.

The general purpose of the Belgian officer selection system is to define to what extent a candidate will be capable of leading a group of people in difficult and dangerous circumstances. Specific training on aptitudes necessary to become a successful military leader is given in a 5 month's officer training period. The specific objective of the psychotechnical assessment is to determine whether the candidate is sufficiently well qualified to pass the training period with success. It is the aim of this study to validate and improve the prediction capacity of this psychotechnical assessment.

Description of the selection procedure.

The person characteristics judged important for success in the training period are : intelligence, decisiveness, initiative, social adaptability, dynamism, influence, sense of responsability, presentation, motivation and physical fitness. These characteristics are assessed in a two day selection procedure. Intelligence is measured by means of a broad spectrum of standardised intelligence tests (verbal reasoning, reading comprehension, mental labyrinth, logical reasoning, spatial memory, learning-speed, fieldmap memory, and organizing ability). A general level of intelligence is based on a weighted sum of the testscores. In this intuitive ponderation, testscores for reading comprehension and verbal reasoning are double weighted. Trained observers give marks for decisiveness, initiative, dynamism, social adaptability and influence, based on the behaviour of the applicants in 5 small-group tasks without an appointed leader. After studying an autobiographical report, the results of personality tests (Achievement Motivation, Social Anxiety, 16 PF, Lpc), a motivational inventory and a written report on group behaviour observation, the psychologist interviews the candidate. Marks for sense of responsability, presentation and motivation are then given, based on the total available information. A short physical test (± 10') of speed, skill, suppleness balance and strength leads to the mark 'physical fitness'. A final mark is worked out at the meeting of the selection board. This final mark is the only score taken into account in the acceptance/refusal decision.

STUDY 1 : VALIDATION OF THE ACTUAL SELECTION PROCEDURE.

The psychological selection of applicants is evaluated against the criterion "success in the officer training". First, the validity of the final selection mark is examined; second the relation of the different predictors with the criterion is examined, in order to evaluate their prediction capacity.

A. Subjects

Subjects are 3.226 Dutch- and Frenchspeaking reserve officer candidates, having passed the selection procedure from 1975 to 1979. The educational level varies from high school (36 %) and higher education (28 %) to university (36 %). The age of the applicants varies between 19 and 24 years, according to their educational level. The specific composition of this group reflects three main army aspects namely combat (Infantry and Tanks), technical (Artillery and Engineers), and clerical functions (Logistics and Administration).

B. Method

Validity coefficients are calculated between the final selection mark and the criterionvariables. As the training staff is unaware of the selectionscores of the trainees, echoeffects in awarding the criterionmarks are avoided. Criterionvariables are :

a. a score 'occupational knowledge', measured by means of theoretical and practical examinations on subjects s.a. map reading, armament, tactical use of weapons, administration, leadership tasks, etc...,
b. a score 'personality', based on the judgement by the trainingstaff of the personal qualities mentioned before,
c. a score 'physical achievement', based on the performance of the trainees in several physical tests (obstacle race, cross-country race, etc...),
d. a final criterionmark (C-Mark), which is a weighted sum of a, b and c (.7(a)+.2(b)+.1(c)).

C. RESULTS AND DISCUSSION

TABLE 1 : VALIDITY OF THE FINAL SELECTION MARK AND OF THE PHYSICAL FITNESS TEST. (Coefficients corrected for restriction of range)

		FINAL SELECTION MARK			PHY FIT TEST
		OCCUP KNOW	PERS	FIN C-MARK	PHY ACH
INFANTRY	(n=488)	.37	.32	.37	.41
TANKS	(n=390)	.17	.28	.15	.48
ARTILLERY	(n=288)	.31	.36	.35	.40
ENGINEERS	(n=228)	.29	.32	.26	.41
LOGISTICS	(n=530)	.45	.18	.47	.42
ADMINISTRATION	(n= 85)	.38	.11(1)	.42	.42
MEDIAN CORRELATION		.34	.30	.36	.41
TOTAL POPULATION	(n=2009)	.31	.25	.31	.36

(1) All coefficients are significant at $p < .01$, except this one.

Table 1 gives the validitycoefficients of the final selection mark and of the physical fitness test.

Table 2 gives the correlationcoefficients between the selectionvariables and the criterionvariables "occupational knowledge", "personality" and "final criterion mark". The result in table 1 shows that the prediction power of the final selection mark is approximately the same for the three criterionvariables; the validity is about .36. Table 2 shows that for the criteria "occupational knowledge" and "final training mark", the predictor capacity of intelligence is higher than the validitycoefficient of the final selection mark. This suggests that the prediction capacity of the selectionsystem can be improved.

TABLE 2 : CORRELATIONCOEFFICIENTS BETWEEN THE SELECTIONVARIABLES AND THE CRITERIONVARIABLES

		FINAL MARK	BEHAVIOUR IN GROUP						INTERVIEW			INTEL
			DEC	INIT	SOC AD	DYN	INFL	G	RESP	PRES	MOTIV	
OCCUPATIONAL KNOWLEDGE	r	.24	.14	.16	.13	.12	.16	.18	.23	.13	.12	.43
	r'	.31	.18	.20	.16	.14	.20	.21	.30	.17	.15	.45
	r'med	.34	.23	.24	.18	.14	.24	.21	.34	.19	.21	.45
PERSONALITY	r	.19	.14	.16	.12	.14	.15	.12	.15	.15	.14	.12
	r'	.25	.18	.20	.15	.17	.18	.14	.19	.18	.18	.13
	r'med	.30	.19	.18	.15	.14	.19	.15	.20	.15	.14	.20
FINAL SCORE	r	.23	.17	.18	.13	.13	.18	.19	.23	.15	.13	.36
	r'	.31	.21	.23	.16	.17	.23	.22	.29	.18	.17	.38
	r'med	.36	.20	.21	.15	.15	.22	.22	.25	.16	.16	.41

r = not corrected / r' = corrected for restriction of range / r'med = median corrected correlation

STUDY 2 : IMPROVING THE PREDICTION CAPACITY OF THE SELECTION SYSTEM.

The prediction capacity of the selection system can be improved by determining the optimal weight of the most powerful predictors. This optimal selection and ponderation is found in a multiple regression analysis.

A. Subjects

Group one : cfr study one. Group two : Subjects are 261 Dutch- and Frenchspeaking reserve officer trainees, accepted in the training school in 1980-1981. This sample is representative for the total population of accepted candidates in 1980-1981 and highly comparable to group one.

B. Method

The objective of the officer selection system is to assess the most essential officer qualities of a candidate without respect to the specific demands of the different arms. This approach is based on the conviction that the task of an officer is essentially the same in all arms. Besides, the acceptance rules make it impossible to construct a specific prediction formula for each single arm, for a candidate is assigned to an arm after being accepted as a potential officer. In fact the similar hierarchy in the main predictorvariables across arms suggests that such a common prediction formula is possible. In order to reduce the number of variables used in the multiple regression equation, it was decided to reduce the number of 'group behaviour predictors' to one. As correlations between the group behaviour predictors ranged from .81 to .93 (median=.88) (cfr table 3), it is clear that these predictors are in fact one undifferentiated global measure of 'behaviour in group'. A multiple regression analysis is performed to examine whether a weighted sum of the five group behaviour variables exceeds substantially the prediction power of 'influence' or 'initiative'. The resulting regression equation is crossvalidated on group 2. A final multiple regression is then performed on the variables 'intelligence', 'sense of responsability', 'presentation', 'motivation' and 'influence' (or its substitute). Occupational knowledge is taken as the criterion, as it is the most important criterionvariable (it determines 92 % of all criterionvariance). Before computing the regression equation the correlations with the criterion are corrected for restriction of range. Indeed the Officer Selection Board pays more attention to the moral and leadership qualities of a candidate than to his intellectual and physical abilities.

Consequently, the final mark reflects the former characteristics more than the latter and the restriction of range is more pronounced for 'sense of responsability', 'influence', 'presentation', and 'motivation' than for 'intelligence' and 'physical fitness'.

C. Results and discussion

1. The regression equation of the group behaviour predictors :

In a stepwise regression (1) the variables influence, initiative and dynamism were selected, resulting in a multiple R = .1996, which is a better prediction of the criterion 'occupational knowledge' than 'influence' or 'initiative' alone (r = .1641). The standard regression equation for the group behaviour is :

Y'_G = .256 (Influence) + .215 (Initiative) - .318 (Dynamism)

The adjusted multiple correlation R_{adj} = .1970. Crossvalidation of the regression equation on group 2 results in a multiple R = .1603. Although the R of the combined group behaviour predictors is higher than the prediction capacity of each single group behaviour predictor, it loses its gain after crossvalidation.

2. The regression of the variables 'intelligence, sense of responsability, presentation, motivation and G', on 'occupational knowledge :

Table 2 gives a summary of the raw, corrected and median corrected correlations between the selectionvariables and the criteria used in the multiple regression analysis. Table 3 gives the correlation between selectionvariables. Physical fitness, being a specific predictor, is not included in this analysis.

Table 4 gives the results of stepwise multiple regressions on the raw, corrected and median corrected correlationcoefficients.

TABLE 3 : CORRELATIONS BETWEEN PREDICTORVARIABLES (ACCEPTED + NON-ACCEPTED APPLICANTS, n = 3.226)

	BEHAVIOUR IN GROUP						INTERVIEW				
	1	2	3	4	5	6	7	8	9	10	11
1. DECISIVENESS	-	.92	.81	.90	.92	.73	76	.69	.71	.33	.88
2. INITIATIVE		-	.84	.92	.93	.81	.77	.69	.73	.32	.90
3. SOCIAL ADAPTABILITY			-	.81	.83	.67	.75	.68	.70	.28	.85
4. DYNAMISM				-	.92	.58	.71	.65	.72	.29	.88
5. INFLUENCE					-	.82	.76	.71	.71	.33	.90
6. G						-	.68	.62	.56	.31	.73
7. SENSE OF RESPONSABILITY							-	.67	.78	.36	.87
8. PRESENTATION								-	.62	.28	.74
9. MOTIVATION									-	.26	.83
10. INTELLIGENCE										-	.38
11. FINAL SELECTION MARK											-

TABLE 4 : STEPWISE MULTIPLE REGRESSION ON RAW, CORRECTED AND MEDIAN CORRECTED CORRELATIONCOEFFICIENTS

STEP	VARIABLES IN EQUATION	RAW			CORRECTED			MEDIAN CORRECTED		
		R	R^2	INC R^2	R	R^2	INC R^2	R	R^2	INC R^2
1	INTEL.	.4344	.1887	.1887	.4530	.2052	.2052	.4528	.2050	.2050
2	INTEL. RESP.	.4538	.2059	.0172	.4699	.2208	.0156	.4911	.2412	.0362
3	INTEL. RESP. G	.4546	.2067	.0008	.4753	.2259	.0051	.4950	.2450	.0038

(1) DIXON, W.J., 1981

In each case (raw, corrected or median corrected correlation) the contribution of the G-variable is neglectable. The beta-coefficients for the G-variable will even be lower considering the shrinkage after crossvalidation. Besides, the intercorrelations between G and the variables already in the equation are rather high (.68 and .31 with responsability and intelligence respectively). As a result a prediction formula with only the variables intelligence and responsability is chosen. The three regression formula for respectively raw, corrected and median corrected correlations are : 1)Y'=.3994(Intelligence)+.1359(Responsability).
2)Y'=.3986(Intelligence)+.1362(Responsability).
3)Y'=.3797(Intelligence)+.2037(Responsability).

3. Crossvalidation :

Formula 2 is chosen and crossvalidated on group two. The shrinkage of the multiple R is about 8 % (R=.4699 → r=.4322), which is a pretty good result, especially when the small modifications in the selection procedure during the years 1979-1980-1981 (additional tests, personality questionnaires, changements in grouptasks, etc...) are taken into account. The data for occupational knowledge show little variability in the predictor accuracy between arms. However for the final criterionmark the differences are more pronounced. Pearson-correlations between predicted and observed final criterionmarks are higher for clerical (.49) than for technical (.38) and combat (.35) functions. To estimate the capacity of this formula to identify the early training drop-outs, (only drop-outs for characterial and intellectual reasons are taken into account), a two-way frequency table is formed and measures of association are calculated. Table 5 gives, for each interval of the predicted criterionscore, the frequency and proportion of drop-outs and other trainees.

TABLE 5 : OBSERVED FREQUENCIES AND PROPORTIONS OF DROP-OUTS AND OTHER TRAINEES FOR 5 CLASSES OF PREDICTED CRITERIONSCORES

	PREDICTED CRITERIONSCORE											
	LOW		LOW AV.		AVERAGE		HIGH AV.		HIGH		TOTAL	
	f	p	f	p	f	p	f	p	f	p	f	p
DROP-OUTS	66	.24	72	.22	142	.12	32	.07	21	.04	333	.12
TRAINEES WITH FINAL C-MARK	208	.76	249	.78	1010	.88	440	.93	489	.96	2396	.88
TOTAL (1)	274	1.0	321	1.0	1152	1.0	472	1.0	510	1.0	2729	1.0

(1) In this total 150 paratroopers are included.

The data show a dependency of the proportion of successful candidates on the predicted criterionscore. For candidates with low predicted criterionscore (10 %) the proportion of drop-out is .24. For the subjects with high criterionscore (19 %) the proportion of drop-out has decreased to .04 while for the total group of accepted candidates, about 12 % do not finish the training period. A X^2-value of 111.5 with 4 df ($p < .001$) confirms this observed relation between predicted criterionscore, and probability of success/failure.

4. Additional attempts at refining the regression formula :

First, it was examined whether a new weighting of the intelligence tests should improve the predictionpower of the factor intelligence. The analysis performed on group 2 yielded a multiple regression equation with four tests.

Compared to the intuitively weighted sum of all seven tests, this combined score yielded a higher muliple R (.4344 → .4540). Unfortunately the crossvalidation on a thirth group was negative. Second, it was tried whether a new predictor could be based on a number of personality tests. But, also this attempt was unsuccessful for the same reason.

D.Conclusion

Study two proved that it is possible to reduce the number of selection-variables to two (intelligence and responsability) resulting in an increase in selection efficiency (reduction of selection time) while at the same time increasing the selection accuracy. A new final mark composed of about 75 % intelligence and 25 % responsability will predict at least 20 % of the criterion variance. The formula is psychologically meaningful since intelligence and sense of responsability can be considered crucial for military leaderschip.

In the light of these findings it is utmost surprising to see that the selection board pays little attention to the most important predictor (intelligence). The selection board stresses personality factors more than intelligence, because the former are believed to have greater importance in later occupation than in the officer training period.

REFERENCES

DIXON, W.J. (Ed.), BMD : Biomedical computer programs. BERKELEY : University of CALIFORNIA Press, 1981.

KERLINGER, F.N., & PEDHAZUR, E.J., Multiple Regression in Behavioral Research. Holt, Rinehart and Winston, INC., NEW YORK, 1973.

USING OCCUPATIONAL SURVEYS TO DEVELOP AIR FORCE SPECIALTY TRAINING STANDARDS

FREDERICK B. BOWER, JR., MAJOR, USAF
Detachment 805 AFROTC, Texas A&M University

As the need for effective management of our military budget becomes more critical each year, we must constantly seek out effective ways to utilize both personnel and material resources. With this in mind, Air Force Air Training Command (ATC) training managers are developing programs which reduce the amount of resident training airmen receive. These reduced resident training programs result in lower training costs and higher productivity by placing airmen on the job sooner. However, such a course must be carefully developed. In order for a program of this type to work, training cannot be reduced arbitrarily. Airmen must be trained to perform those duties and tasks required in their first job or first enlistment.

All ATC curriculum managers have been tasked with structuring basic airmen resident courses to prepare students for their first job. This entails not only a review of each resident course but also a review of the entire airmen training program. This is accomplished through careful analysis of a career's Specialty Training Standard (STS) which serves as the primary training outline for a career ladder. STS requirements identify needed training for the first job. Many STS requirements not required in the performance of the first job are now taught on-the-job (OJT). Other more technical STS requirements will be in follow-on resident training courses after the first enlistment. Consequently only career motivated individuals receive the extensive resident training required to perform the first job.

Beginning in 1977, at the direction of the ATC Commander, General John W. Roberts (1977), utilization and training workshops (initially called course scrubdowns) were scheduled to provide a forum for discussion of training needs. Training Managers, and users of the trained product all provide in-put toward course development. However, participants in the workshops had no means of displaying reliable data to substantiate training requirements. As a result, participation by the USAF Occupational Measurement Center, Airmen Career Ladders Analysis Section, was included in the workshop agenda. Personnel from this organization, were asked to furnish occupational survey data collected from airmen serving in the specialty being evaluated. The effects of using job analysis data in this way have had a profound and significant effect on the way training is now being reviewed within ATC.

The use of job analysis data in making training decisions is not a new concept. Morsh (1964) described one objective of the Air Force Occupational Survey Program as the determination of training needs. Since that time, the same theme has often been repeated, most recently by Keeth (1977) and Turner (1978). However, until institution of the Utilization Workshops, the full impact of job data as a training tool had never been so fully demonstrated. Other military services and institutions in the civilian sector are also using job analysis to make training decisions as was reported by Davis (1978) and Cunningham and Drewes (1978), but not to the extent now used by ATC. The Air Force, more than any other service, employs a system that outlines an airman's training requirements for a complete career in a single job specialty, and documentation for this system is the STS.

The STS as described by Air Force Regulation 8-13, Air Force Specialty Training Standards, outlines the training required achieve various skill levels

within an enlisted Air Force Specialty (AFS). Through its use the individual training of airmen is standardized and the quality of training controlled. STSs are designed to perform the following:

a. Describe tasks, knowledges, and proficiency level requirements for one of more AFSs.

b. Specify the degree of training provided in formal schools.

c. Identify career development courses (CDCs) and additional references needed for upgrade and qualification training, and serve as a review for specialty knowledge tests (SKTs). As such, they are used:

(1) As course specification documents.

(2) For basic reference by major air commands in evaluating course graduates.

(3) As the basis for preparing career development courses.

(4) As a guide for establishing local OJT programs.

(5) As the basis for development of SKTs.

It is easy to see that development of STSs using occupational survey data encompasses more than just a basic training course, because the STS influences an individual's total training program and his job classification and promotion testing opportunities as well.

Flournoy (1978) traced the evolution of the STS form the earliest requirement for documentation of OJT to its present form. Air Force managers recognized the need for standardized training outlines in order to insure that airmen were trained to perform the job they were assigned. The increased size of the peace time Air Force, a rapid turnover of experienced personnel, and the constant increase in the cost of formal training initiated the movement away from formal training and toward the documented OJT program we have today. While this paper is not intended to debate the advantages and deficiencies of the STS it should be pointed out that the STS is the only single document currently being used by the U.S. military services that lists tasks, knowledges, and the skill levels required of an individual to progress satisfactorily in a chosen profession through a complete career.

The STS consists of two primary sections; the tasks, knowledges and study references section, and the proficiency level, progress record, and certification section. The tasks and knowledges are listed in columnar fashion with their associated study references. To the right of each task or knowledge element is the proficiency level and space for recording the progress and certification of that element for each of the three technical skill level progressions. The 3 -, 5 - and 7 - skill levels equate to the apprentice, specialist, and technician level of competence. The proficiency level code may be a task performance task knowledge, or subject knowledge level. These codes are explained in Figure 1.

Until ATC began the Utilization and Training Workshops, the tasks, knowledges and proficiency levels listed on STS documents were, for the most part, determined by subject matter experts assigned to ATC or through in-puts submitted by technicians in the field. Although occupational survey data is routinely sent to those responsible for STS construction, training managers often were not trained to extract the needed information from all the data available. Now, however, occupational survey analysts are present at the workshops to provide that service. Therefore occupational survey results have been most effective in enhancing the development or review of STS documents.

The process is relatively simple. Subject matter experts are asked to match each task statement in a job inventory for a given specialty to an item in that specialty's STS. Occupational survey analysts can then evaluate the STS in three ways.

An STS item is first looked at in terms of what percentage of the personnel report using tasks identified with that item. This evaluation may be by skill level, time in service, (rank or other identifiable grouping). Although percent of members performing is not a criteria for inclusion or deletion of an item from the STS, a criteria does exist for inclusion or deletion of items taught in formal resident training. ATC Regulation 55-22, Occupational Survey Program, sets a minimum criteria to be applied in design or revision of basic resident training courses of 30 percent of first job/first enlistment airmen performing any given task in a job inventory. For tasks where the probability of performance by this group is less than 30 percent, resident training is not recommended unless such training can be justified (as for safety reasons). Therefore, all subject matter areas covered in a resident training course will be listed in the STS, but all STS items need not be covered by formal resident training.

The fact that a task statement elicits a high response rate, however, does not mean that the task must be listed in the STS. Analysts next look at each task in relation to its task learning difficulty rating. Task learning difficulty is a secondary factor routinely collected during the occupational survey administration. Briefly, experienced senior airmen in the specialty being surveyed are asked to rate each task in the job inventory based on the time it would normally take an airman in the specialty to learn to do the task. Ratings are from one (very small amount of time) to nine (very large amount of time). Combined ratings are then standardized so that a rating of five represents an average amount of time spent to learn a task. The development and validation of task learning difficulty is explained in more detail by Mial and Christel (1974). By comparing task difficulty rating to the task statements and the percent of members performing the tasks, it is easier to see just what is required in terms of OJT and/or formal training. Obviously tasks with low difficulty ratings may require little or no formal training, and for that reason they may also have no need for being listed in the STS.

The third way of evaluating the STS is through the use of the training emphasis rating. As reported by Ruck, Thompson, and Thomson (1978), training emphasis is a secondary factor collected in the manner of task learning difficulty ratings. The difference is that subject matter specialists are asked to rate each task statement on a nine point scare (extremely litte to extremely heavy) in terms of whether formal training (school or OJT) should be emphasized for first enlistment airmen. This data can be used to cross reference tasks with high response rates or high task difficulty ratings in order to justify formal training and inclusion in the STS. They may also justify formal training or inclusion in the STS of inventory tasks with low response rates if subject matter experts in the field believe them to be important.

What of tasks with low responses and low task learing difficulty or training emphasis rating, but are unique and important to a specific agency or unit within the specialty population? The Air Force has provided for this situation through the use of the Air Force Form 797, Job Proficiency Guide (JPG). The JPG is used to document training required of an individual above the normal requirements for a given specialty. The JPG is prepared by the agency requiring

additional training and is attached to the STS by the unit providing the training. In this manner, the STS remains a general document listing only training required by most airmen assigned to the specialty. Thus, unnecessary training is precluded, but the capability to identify and document additional requirements is available when needed. A full discussion of the JPG can be found in Air Force Regulation 50-23, On-The-Job-Training.

In order to effectively utilize the survey data, a computer product developed by Thew and Weissmuller (1978), the modular factor printout, is being used by occupational analysts and provided to training managers. As shown in Figure 2, the tasks in the job inventory are clustered under their corresponding STS item. The training emphasis rating, task learning difficulty rating, and the percent of members responding by skill level are displayed to the right of each task statement. In this single printout, all the survey data used to make a training decision are displayed for each task. Although the printout is time consuming and expensive to produce, the data is presented in a manner that is comprehensive and understood by decision makers not generally accustomed to using computer generated products.

The impression should not be left that occupational survey data alone could be used to revise or develop training documents or formal training programs, rather the data is another tool for training managers and subject matter experts to use and weigh in relationship to other factors. The quality and completeness of the job inventory, and the timeliness of the survey bear on the usefullness of the data. Training costs, system procurement, programmed changes in personnel utilization, and equipment modification all must be considered when determining whether tasks can or should be trained. Job analysis is just a part of the Instructional System Development (ISD) model used by the Air Force for designing training programs.

The point to be made is that unlike other methods of employing job analysis to define and design training, the Air Force method relies on a cross-check approach of evaluation of an established training outline (STS) encompassing both formal resident training and OJT, rather than starting from the beginning each time training is reviewed. This method allows for use of occupational data to be applied to the identification of training needs beyond the classroom without creating redundancy of training, because both the technical training centers and field trainers are following and documenting training on the same outline.

How successful has the Air Force been in developing effective training programs while reducing costs? Figures from just one training center reported by Meece (1979) reveals that savings have been significant and course graduates are reporting to the field better prepared to perform their assigned first job. While the use of occupational survey data to revise and develop STSs cannot be credited for all of the savings, reports from training managers indicated that such savings would never have been achieved had the job survey data not been employed. As a result, Headquarters ATC, Technical Training, has formalized a system of scheduling workshops to coincide with the completion of occupational survey reports on some career specialties needing a review of training requirements. This system has been included in both Air Force and ATC regulations to institutionalize the systems. The combination of an integrated scheduling of workshops and improvements of occupational survey data for use by curriculum developers and workshop participants suggests that the Air Force will continue to enhance its management of our very critial training dollars.

BIBLIOGRAPHY

AFR 8-13, Air Force Specialty Training Standards; Washington D.C. Hq AF, June 1974.

AFR 50-23, On-The-Job-Training, Washington D.C., HQ AF, 29 May 1979.

ATCR 52-22, Occupational Survey Program; Randolph AFB, HQ ATC; 22 August 1978.

Davis, D.D.; "Data Base to Determination of Training Content: A Managable Solution"; Symposium Papers (20th Annual Conference of the Military Testing Association, Oklahoma City, Oklahoma; 30 October - 3 November 1978, Vol 1, 28-50).

Flournoy, D.B.; Air Command and Staff College Student Research Report"; An Evaluation of the Air Force Specialty Training Standard (STS)"; Maxwell AFB, Alabama, 1978, 055-78.

Keeth, J.B.; "The USAF Occupational Survey Program"; Symposium Paper (19th Annual Military Testing Association Convention, San Antonio, Texas, 17-21 October 1977) Lackland Air Force Base Texas; USAF Occupational Measurement Center, 77-14, December 1977, 12-16.

Letter, Subject: Expanded Course Scrubdown, 1 Nov 77 from General John W. Roberts, CC ATC to all ATC Technical Training Center Commanders.

Meece, C.C., "Training Management and Utilization Workshops (Scrubdowns)"; Report to Commander; Keesler Technical Training Center, Keesler AFB, Mississippi; January 1979

Mial, R.P., Christel, R.E.; "The Determination of Training Priority for Vocational Tasks"; Preceedings, Psychology in the Air Force Symposium; USAF Academy, Colorado April 1974.

Morsh, J.E.: "Job Analysis In the United States Air Force"; Personnel Psychology; 1964; 17, 7-17.

Ruck, H.W., Thompson, N.A., Thomson D.C.; "The Collections and Prediction of Training Emphasis Ratings for Curriculum Development"; Symposium Papers (20th Annual Conference of the Military Testing Association, Oklahoma City, Oklahoma; 30 October - 3 November 1978, Vol 1, 242-257).

Thew, M.C., Weissmuller, J.J.; "CODAP' A New Modular Approach to Occupational Analysis"; Symposium Papers (20th Annual Conference of the Military Testing Association, Oklahoma City, Oklahoma; 30 October-3 November 1978, Vol 1, 362-371)

Turner, J.A.; "The USAF Occupational Survey Program - An Aid to Force Management"; Symposium Papers (6th Annual Psychology in the DOD Symposium, USAF Academy, Colorado, 22 April 1978). Lackland Air Force Base, Texas; USAF Occupational Measurement Center Technical Note Series 78-01, April 1978, 1-4.

Figure 1

PROFICIENCY CODE KEY		
	SCALE VALUE	DEFINITION. The Individual
TASK PERFORMANCE LEVELS	1	Can do simple parts of the task. Needs to be told or shown how to do most of the task (EXTREMELY LIMITED)
	2	Can do most parts of the task. Needs help only on hardest parts. May not meet local demands for speed or accuracy (PARTIALLY PROFICIENT)
	3	Can do all parts of the task. Needs only a spot check of completed work. Meets minimum local demands for speed and accuracy. (COMPETENT)
	4	Can do the complete task quickly and accurately. Can tell or show others how to do the task. (HIGHLY PROFICIENT)
* TASK KNOWLEDGE LEVELS	a	Can name parts, tools, and simple facts about the task (NOMENCLATURE)
	b	Can determine step by step procedures for doing the task. (PROCEDURES)
	c	Can explain why and when the task must be done and why each step is needed. (OPERATING PRINCIPLES)
	d	Can predict, identify, and resolve problems about the task. (COMPLETE THEORY)
** SUBJECT KNOWLEDGE LEVELS	A	Can identify basic facts and terms about the subject. (FACTS)
	B	Can explain relationship of basic facts and state general principles about the subject. (PRINCIPLES)
	C	Can analyze facts and principles and draw conclusions about the subject. (ANALYSIS)
	D	Can evaluate conditions and make proper decisions about the subject. (EVALUATION)

– EXPLANATIONS –

* A task knowledge scale value may be used alone or with a task performance scale value to define a level of knowledge for a specific task. (Examples: b and 1b)

** A subject knowledge scale value is used alone to define a level of knowledge for a subject not directly related to any specific task, or for a subject common to several tasks.

– This mark is used alone instead of a scale value to show that no proficiency training is provided in the course, or that no proficiency is required at this skill level.

X This mark is used alone in course columns to show that training is not given due to limitations in resources.

Figure 2

906X0 - Medical Administrative STS Analysis FCPRT1 PAGE 12

D TSK	TITLES	TNG EMP *D*	TSK DIF (F)	906 30 (M)	1ST JOB (M)	906 50 (M)
N 432	Annotate Alpha Rosters with incoming or outgoing personnel information	2.83	3.78	13.9	12.8	8.7
086 16D.	Eligibility for Medical Care					
F242	Determine Admission Elgibility	5.96	4.44	20.5	11.9	15.6
N460	Verify Identification of Patients	5.17	3.99	37.3	35.3	20.7
G283	Verify Eligibility of Air Force Reserve Admissions to Hospital	4.55	4.88	6.8	7.8	9.1
G285	Verify Eligibility of Civil Service Employee Admission to the Hospital	4.53	5.00	6.3	7.3	8.0

SUBJECTIVE APPRAISAL AS A FEEDBACK TOOL

Billy L. Burnside

U. S. Army Research Institute[1]
Fort Knox Field Unit

Start

The products of U. S. Army Centers/Schools are trained graduates and training support materials. In order to appraise the quality and utility of these products, training developers and evaluators in the Centers/Schools need meaningful feedback from users at the institution and in the field. There are six principle methods which these personnel may use to obtain such feedback: receipt of informal comments, administration of surveys/questionnaires, conduct of interviews, analysis of available unit performance records, observation of training classes and exercises, and administration of performance tests. Interviews with battalion commanders and staffs (Burnside, 1981) and with training developers and evaluators in a typical Center/School (Witmer and Burnside, 1982) indicate that the first three of these methods are the most commonly used. A common attribute of these three methods is that they are relatively subjective in nature; i.e., they are largely based upon individuals' perceptions, judgments, and opinions.

Since the feedback presently available to training developers and evaluators consists largely of subjective data, an important issue to be addressed is how accurate or valid these data are. That is, how do they compare with data gathered using more objective methods and criteria? This issue is addressed in the present paper by reviewing research results comparing subjective ratings gathered using surveys or interviews with relatively objective data gathered using structured observations or "hands-on" performance tests. The type of feedback of interest here is appraisal of the performance of individual soldiers and military units on specific tasks, rather than assessment of general knowledge and abilities. An example of subjective appraisal is using a survey or interview to ask a soldier whether he or she can perform a specific task. The comparable objective appraisal would involve administration of a "hands-on" test in which the soldier's performance was compared to a validated standard. Subjective appraisal is a relatively efficient and cost-effective method of gathering feedback, so it will continue to be used in the military. The key question thus becomes whether the data gathered using this approach are sufficiently accurate to warrant their use in particular situation, and whether their accuracy can be increased by refinements in collection methodologies.

The aspects of subjective feedback addressed in this paper include what is appraised, who does the appraising, and how the appraisal is done. The type of appraisal of greatest interest here involves estimates of soldiers' proficiencies on specific tasks, but other types addressed include judgments of the criticality, difficulty, and performance frequency of specific tasks.

---

[1]The views expressed in this paper are those of the author and do not necessarily reflect the view of the U. S. Army Research Institute or the Department of the Army.

These are the types of estimates typically obtained using Comprehensive Occupational Data Analysis Program (CODAP) surveys. The issue of who does the appraisal is addressed by summarizing research relating to self-appraisals, supervisory appraisals, and peer appraisals. Discussion of the issue of how subjective appraisals are done centers around survey and interview techniques, and the paper concludes with discussion of ways to improve the accuracy of subjective data.

Types of Appraisals

Proficiency

A key element of feedback to Army Centers/Schools is data relating to the proficiency with which soldiers can perform specific required tasks. Such data are needed to allow training developers to evaluate both institutional and unit training and to make modifications as needed. Since the operational testing of soldiers' performance is costly in terms of time and resources, proficiency data are usually gathered through subjective estimates. That is, soldiers are asked to estimate their confidence or the likelihood that they can perform specific tasks. Supervisors may also be asked to rate soldiers' proficiencies. How accurately do such subjective appraisals reflect actual task proficiencies?

Several pieces of research conducted outside the military are relevant to answering this question. There is some evidence that people can appraise their own task-specific proficiencies with moderate accuracy, as long as the tasks appraised are basic ones with which they have had extensive experience. For example, Ash (1980) found that self-ratings of straight copy typing ability correlated in the .44 to .59 range with the results of typing tests. However, subjective ratings of more complex typing skills did not correlate as highly with performance. In a recent meta-analysis of self-evaluation of ability, Mabe and West (1982) found the mean correlation between self-evaluation and performance measures to be approximately .30. While they found many methodological weaknesses that limited the interpretation of correlational data, the general conclusion is that self-appraisals of proficiency are not particularly accurate. In a meta-analysis of educational research, Cohen (1981) found that the mean correlations between students' subjective appraisals of instruction and measures of students' proficiencies ranged from .38 to .47. He also identified several methodological problems, such as the lack of objective criteria to compare subjective appraisals against and the fact that most appraisals obtained have been global rather than task-specific in nature. DeNisi and Shaw (1977) avoided some of the common methodological problems by examining the accuracy of self-appraisals for specific abilities on tasks such as visual pursuit and manual speed and accuracy. While the correlations between self-appraised and tested abilities were almost all statistically significant (in the .20 to .40 range), they showed that these results had little practical significance. Due to methodological weaknesses in the relevant research and problems in interpreting correlations in the .30 to .40 range, the appropriate conclusion appears to be that there is no convincing evidence that subjective appraisals of proficiency are accurate.

Few studies of the accuracy of subjective appraisals of proficiency have been conducted in a military setting. Many of those that have been conducted

have suffered from methodological problems, such as the lack of objective criteria or the lack of specificity or explicitness in the tasks addressed. For example, Hall, Denton, and Zajkowski (1978) found that supervisors' estimates of sailors' proficiencies on several tasks did not correlate significantly with performance. However, the criterion used was performance on a written test rather than "hands-on" performance. A further examination of two sets of data previously published by the Army Research Institute provides some insights that have not previously been available.

Hiller (1980) collected data which allow comparison of self-estimates and "hands-on" performance test results for five specific tasks. The general finding is that self-appraisals of proficiency were accurate for general leadership skills, were at best moderately accurate for cognitive skills, and were inaccurate for motor skills. The accuracy of subjective appraisals was thus found to decline as the objectivity of the performance test criterion and standards increased. Leadership skills are difficult to develop standards for and objectively evaluate; the high accuracy of self-appraisals of leadership skills may have resulted from the comparison of these appraisals with results of relatively subjective performance tests. Relatively objective performance tests are available for "hands-on" motor skills, and self-appraisals of such skills were highly inaccurate. This indicates that subjective appraisals of proficiency are not accurate when compared to an objective criterion.

In the military skill retention literature, several instances can be found in which self-appraisals of proficiency were collected prior to a retention test, but the results were not reported. This leads one to suspect that the results were negative; i.e., that the self-appraisals were not found to be accurate. This suspicion is supported by further examination of data collected by Shields, Goldberg, and Dressel (1979), in which confidence ratings of proficiency on 20 tasks were found not to significantly correlate with performance test results. It thus appears that retention research has not supported the accuracy of subjective appraisals of proficiency.

The data reviewed above indicate that subjective appraisals of proficiencies (largely in terms of self-appraisals) on specific tasks often do not represent true abilities. This appears to be especially true when the subjective appraisals are compared to objective well-specified performance criteria. Before subjective appraisals are used as feedback to training developers, the relationship between such appraisals and more objective measures of performance should be further examined. Self-ratings of proficiency may only be accurate when addressing explicit tasks with which the ratees have extensive experience.

Criticality

Since training resources are limited, training developers must somehow determine which tasks are most critical for combat performance and therefore most important to train. This is typically accomplished by preparing an extensive list of tasks and asking subject matter experts to subjectively rate their criticality. Just as with estimates of proficiency, one can question how accurately subjective appraisals of criticality represent the "true" relative importance of tasks. Data are relatively sparse in this area, but those available indicate that rater agreement (interrater reliability) has generally been found to be low. The accuracy or predictive validity thus would be

expected to be low. Another problem in this area is the specification of an objective criterion of criticality. Due to these reliability and criterion problems, subjective appraisals of task criticality should be used cautiously, if at all.

Difficulty

Knowledge of the relative difficulty of tasks is important to training developers, in order to determine the proper distribution of training time and resources. Appraisals of task difficulty are usually made subjectively, based upon the experiences and opinions of subject matter experts. Indications are that subjective appraisals of task difficulty are not generally accurate; i.e., the tasks picked as most difficult by subject matter experts are not the ones most commonly failed by soldiers. Part of the reason for this problem may lie in the fact that difficulty is not consistently defined. Some tasks are difficult to learn but not to perform, and vice versa. Raters having different perceptions of what is meant by difficulty would thus provide unreliable ratings for such tasks. In obtaining subjective appraisals of task difficulty, care must be taken to precisely define the rating dimension.

Frequency

While limited relevant data are available, indications are that judgments of the frequency with which specific tasks are performed are not generally accurate. Again, there is a criterion problem here, since objective measures of task performance frequency can only be obtained through laborious observation in the field. In cases where this has been done (e.g., Johnson, Tokunaga, and Hiller, 1980), accurate frequency estimates have been obtained only for broad categories of tasks addressed through carefully controlled data collection techniques. As with the other types of subjective appraisal addressed above, frequency estimates should not be assumed to be accurate. They should be collected very carefully and their accuracy should be checked against objective criteria.

Types of Appraisers

A primary consideration in the use of subjective appraisals is the sources from which they are collected. Three general alternative sources are available for providing subjective appraisals as feedback: soldiers evaluating themselves (self-appraisal), supervisors, and peers. Research on the relative accuracy of these appraisal sources has produced mixed results; it is difficult to address the relative accuracy of these sources when the absolute accuracy of each of them is undetermined.

The biggest advantage of self-appraisals is that individuals have extensive data available about themselves and can provide information that is unavailable from other sources. Individuals are aware of situational factors in their own behavior, and are less likely to over-generalize than outside observers are. A problem with self-appraisals is that individuals may not be capable of appraising themselves accurately, as shown by the research summarized in the previous section. Another problem is that individuals may have reason to bias their self-appraisals in a positive direction, resulting in leniency errors. Such errors are common in self-appraisals, but they can be reduced by techniques such as making the appraisals publicly verifiable

(van Rijn, 1981). When self-appraisals are used, their accuracy should be checked against an objective criterion, and the appraisers should be aware that this is being done.

The research literature does not at this time allow any definitive conclusions on the relative accuracy of subjective appraisal sources. What is needed is a study which includes the collection of supervisory, peer, and self-predictions of proficiencies on specific tasks, followed by objective measures of task performance. The literature thus far has generally failed to include objective criteria for comparison purposes, and until it does the accuracy issue will be unresolved. Self-appraisals often suffer from leniency biases, and peer and supervisory appraisals may suffer from tendencies to overgeneralize from small samples of data. Accuracy of these approaches should thus not be assumed, but should be checked against relatively objective criteria.

Types of Appraisal Methods

The previous discussion leads to two primary conclusions about subjective appraisal. The first of these is that adequate data are not yet available to determine either the absolute accuracy of subjective appraisals or the relative accuracy of different appraisal sources. The second is that the limited research which has directly addressed the accuracy of subjective appraisals has in general not found it to be high. These appraisals should thus be used very cautiously with frequent checks on their accuracy. However, military agencies will continue to use subjective appraisals as feedback, due to the ease with which they can be collected. Recognition of this fact leads to the need to identify ways in which the accuracy of subjective appraisals can be increased. A review of the literature by the present author and a meta-analysis reported by Mabe and West (1982) has indicated several ways in which this can be done. These are briefly summarized below.

1. Integrate mutually supportive subjective appraisal methods within a feedback system. Since no appraisal method is complete and sufficient in and of itself, methods should be used to complement each other. Surveys can be used to obtain a general overview of the situation, interviews can be used to obtain more in-depth detail on specific problems, and observations and performance tests can be used as accuracy checks.
2. Ensure that question developers and subjective appraisers have a common basis of understanding. These groups should share a common understanding of task elements, successful task completion, appropriate standards, and rating dimensions.
3. Design questions to maximize accuracy. Make the situation and behavior being addressed as explicit as possible, and specifically state the action being addressed.
4. Make rating scales as explicit as possible. Phrase rating scales in terms of observable measures of performance, rather than in vague, general terms.
5. Be sure that raters have had experience with the tasks rated. Ensure that supervisors have had ample opportunity to observe task performance by the people they are rating.

6. Train raters before they provide subjective appraisals. This training should include experience with the rating scales to be used, a discussion of common types of psychometric errors, and a discussion of the dimensions of the situation being evaluated.

7. Facilitate raters' recall of relevant experiences. Ask raters to review their previous experiences, provide them with thorough descriptions of the tasks and situations being rated, and provide any other cues which aid memory.

8. Make certain that appraisers have the cognitive capacity and motivation to provide accurate ratings. Explain the need for accurate rating data during instructions. Check the accuracy of subjective ratings whenever possible, and let the raters know that this will be done.

References

Ash, R. A. Self-assessments of five types of typing ability. Personnel Psychology, 1980, 33, 273-282.

Burnside, B. L. Field performance feedback - A problem review (Draft Research Report 1323). Alexandria, VA: US Army Research Institute, August 1981.

Cohen, P. A. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 1981, 51, 281-309.

DeNisi, A. S. & Shaw, J. B. Investigation of the uses of self-reports of abilities. Journal of Applied Psychology, 1977, 62, 641-644.

Hall, E. R., Denton, C. F., & Zajkowski, M. M. A comparative assessment of three methods of collecting training feedback information (TAEG Report No. 64). Orlando, FL: Training Analysis and Evaluation Group, December 1978.

Hiller, J. H. A methodology for estimating the cost-effectiveness of alternative pretesting procedures (Draft Technical Report). Alexandria, VA: US Army Research Institute, November 1980.

Johnson, C. A., Tokunaga, H. T., & Hiller, J. H. Validation of a job analysis questionnaire against intensive observation. Paper presented at the Military Testing Association Conference, Toronto, October 1980.

Mabe, P. A. III & West, S. G. Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology, 1982, 67, 280-296.

Shields, J. L., Goldberg, S. L., & Dressel, J. D. Retention of basic soldiering skills (Research Report 1225). Alexandria, VA: US Army Research Institute, September 1979.

van Rijn, P. Self-assessment in personnel selection and placement. Paper presented at the Military Testing Association Conference, Arlington, VA, October 1981.

Witmer, B. G. & Burnside, B. L. Feedback needs of training developers and evaluators (Draft Research Report). Alexandria, VA: US Army Research Institute, September 1982.

AD P00853

IMPLEMENTATION OF AN AIR FORCE OCCUPATIONAL RESEARCH DATA BANK

Roger L. Camp

OAO Corporation
Brooks AFB, Texas 78235

Richard H. Brown
Hendrick W. Ruck

Air Force Human Resources Laboratory
Brooks AFB, Texas 78235

INTRODUCTION

An Air Force Occupational Research Data Bank (ORDB) is being implemented to provide ready access to a variety of current and historical occupational information for research and management use. The prime contractor for this effort is the GSA data processing services contractor, currently OAO Corporation. OAO personnel assigned to this project are co-located with the monitoring activity (AFHRL/MODS) at Brooks AFB, Texas.

Development of the ORDB began in 1978 with the investigation of available data that would contribute to Air Force occupational analysis and management. A number of sources and types of information were identified and have been obtained for inclusion in the ORDB. Major types include hard copy reports and studies, statistical variables summarized for occupations from individual Air Force members and technical training course data, and Comprehensive Occupational Data Analysis Program (CODAP) studies performed at the Air Force Occupational Measurement Center (OMC) and AFHRL (Carpenter, Archer, & Camp, 1979; Stephenson, 1979).

These types of information have been obtained and are being incorporated in the ORDB. The system which provides for storage and on-line retrieval of the information is described in the following section.

SYSTEM OVERVIEW

The ORDB operates on the AFHRL's Univac 1100/81. Five subsystems are tailored to the types of data and kinds of retrieval needed by the user. These subsystems are linked together by a front end program to simplify the use of the ORDB. The programs are designed to interact with the user, assisting in the choice of the appropriate subsystem, and in selecting the desired information. Each subsystem is described below.

1. Computer Assisted Reference Locator (CARL) Subsystem. The CARL system is used to retrieve references to occupationally-related information, such as published studies, technical reports, recurring reports and films. Retrieval is based on user selected keywords. CARL was obtained from the Navy Personnel Research and Development Center (NPRDC) and modified to operate on the 1100/81 (Sands, 1978; Sands & Hartman, 1979). Additional modifications were made to accept Air Force Specialty Codes (AFSCs) as keywords, and to clarify user

selection of output options. Two search techniques have been added to CARL. The first utilizes a binary search to speed the retrieval of references for known keywords. The second uses a string-search process to compare a user-input keyword, such as "JOB" to keywords on file, retrieving keywords "JOB REQUIREMENTS," JOB ANALYSIS" and "JOB DIFFICULTY" and the number of references for each, for example. This enables the user to determine the available keywords most likely to meet requirements, and to structure the inquiry accordingly. Sample output is provided in Figure 1. CARL is written in FORTRAN.

2. Statistical Variable Subsystem. This subsystem provides on-line retrieval of summary statistics for enlisted AFSCs, Career Ladders and Career Fields. Information available in this subsystem includes duty descriptions (Figure 2), prerequisites, and demographic, performance rating, aptitude scores, and training variables. Within specialty, data is organized by enlistment status (1st term, 2d term, Career, Total) and for the accession cohort during a given calendar year. The information reflects the characteristics of the Air Force enlisted personnel, by specialty, for different calendar years (see Figure 3). Data for 1978 and 1979 has been loaded, and that for 1980 and 1981 is being generated. 1982 data will be added after the end of December, 1982. This subsystem uses System 2000 (S2K) Data Base Management System with COBOL extension (Intel Corporation, 1982).

3. CODAP Report Display Subsystem. This subsystem was developed to provide the task scientist and manager with the ability to rapidly retrieve OMC and AFHRL CODAP reports and review them on the terminal screen. To accommodate the standard CODAP report format, Datagraphix 132 character remote terminals are in use at principal user sites. Studies can be selected by either AFSC, study number, or from a menu of available studies. Studies from 1978 to present have been loaded, and any report retrieved on the screen can also be printed at the user's option, as can be seen in Figure 4. This subsystem is programmed in PRISM (AFHRL/TS, 1982).

4. Cross-Study Analysis. This subsystem was developed in response to the need to compare CODAP reports across specialties. Since CODAP variable numbers and titles are not necessarily standard, identifying corresponding data in different studies presented a difficult task. To solve this problem, studies are indexed as they are loaded to the ORDB for a set of 15 variables and 8 groups. The variables include: Number of Tasks, Average Task Difficulty Per Unit Time Spent (ATDPUTS), Job Difficulty Index, Grade, Major Command, Time in Career Field (TICF), Total Active Military Service (TAFMS), Eligible to Reenlist, Eligible for Retirement, Job Interest, Talent Utilization, Training Utilization, Sense of Accomplishment, Plan to Reenlist, and How Assigned to Present Career Field.

Groups that can be analyzed include: Total Sample, Skill Levels 3, 5, 7, 9, 1-48 Months TAFMS of TICF, 49-96 Months TAFMS or TICF, and 97+ Months TAFMS or TICF. On-line retrieval of corresponding data from multiple studies on one or a number of job groups can be performed using this system. For example, job difficulty of airmen with 1-48 months TAFMS or TICF can be retrieved for comparison across any number of AFSCs. This subsystem is written in PRISM and uses the same data files as the CODAP Report Display subsystem. Sample output is provided in Figure 5.

5. Comments. The comments subsystem provides an opportunity for users and developers to record information related to the ORDB while using the remote terminal. Comments can include anything relevant to data contained in the

system, or to the system operation itself. It's been especially useful as a means of obtaining user feedback during the development effort, and for announcing the implementation of enhancements or changes

PROJECTED ENHANCEMENTS

Enhancements projected for the ORDB fall in two main categories: new data and new capabilities. References will be added to the CARL system as new items are published, generated or produced. The ORDB statistical data base will contain five calendar year's data when complete. As a new year is added, the oldest year's data will be saved on tape and deleted from the master file to save space. As CODAP studies are completed, reports will be extracted for addition to the CODAP Report Display file.

New capabilities for the ORDB include an interface with the Statistical Package for the Social Sciences (SPSS), and a batch process for Cross-Study Analysis. The statistical data base will be examined to determine the types of SPSS analysis that would be appropriate to the statistical variables it contains. Procedures will be developed to convert selected S2K data into SPSS-compatible data. It is anticipated that correlation and trend analysis, as well as other SPSS techniques, will prove useful to researchers.

Cross-Study retrieval will be expanded to include more data and format manipulation, with the development of batch programs to extract and compare CODAP data from different studies, and to generate analysis reports that until now have been impracticable.

References

AFHRL/TS. Programming instructions for string manipulation (PRISM). Developed and maintained by Air Force Human Resources Laboratory, Technical Services Division, Brooks AFB, Texas 78235, 1982.

Carpenter, J. B., Archer, W. B., & Camp, R. L., Establishing an Air Force occupational research data bank. Proceedings of the 21st Annual Conference of the Military Testing Association, San Diego, California, October 1979.

Intel Corporation. System 2000 (S2K) data base management system. Intel Corporation, P. O. Box 9968, Austin, Texas 78766, 1982.

Sands, W. A., Computer assisted reference locator (CARL) system: An overview. Proceedings of the 20th Annual Conference of the Military Testing Association, Oklahoma City, Oklahoma, October 1978.

Sands, W. A., & Hartman, L., Research management applications of the computer assisted reference locator (CARL) system. Proceedings of the 21st Annual Conference of the Military Testing Association, San Diego, California, October 1979.

Stephenson, R. W. Development plans for an Air Force occupational research data bank. Proceedings of the 21st Annual Conference of the Military Testing Association, San Diego, California, October 1979.

Figure 1.

ORDB COMPUTER-ASSISTED REFERENCE LOCATOR

KEYTERMS SPECIFIED FOR THIS QUERY:
1. 811XD

REFERENCE 969

CHRISTAL RE 72A
ANALYSIS OF RACIAL DIFFERENCES IN TERMS OF WORK ASSIGNMENTS, JOB INTEREST, AND FELT UTILIZATION OF TALENTS AND TRAINING. AFHRL-TR-72-1, AD-741 758. LACKLAND AFB, TX: PERSONNEL RESEARCH DIVISION, JANUARY 1972. PROJECT 7734. NTIS.

REFERENCE 999

WILEY LN 72A
ANALYSIS OF THE DIFFICULTY OF JOBS PERFORMED BY FIRST-TERM AIRMEN IN 11 CAREER LADDERS. AFHRL-TR-72-6C, AD-757 876. LACKLAND AFB, TX: PERSONNEL RESEARCH DIVISION, JULY 1972. PROJECT 7734. NTIS.

REFERENCE 1026

WILBOURN JM GUINN N 73F
FEASIBILITY OF USING SPECIAL MEASURES IN THE CLASSIFICATION AND ASSIGNMENT OF LOWER MENTAL ABILITY AIRMEN. AFHRL-TR-73-31, AD-777 831. LACKLAND AFB, TX: PESONNEL RESEARCH DIVISION, NOVEMBER 1973. PROJECT 7719. NTIS.

Figure 2.

AFSC DESCRIPTIVE DISPLAY

AFSC: 811XD TITLE: SECURITY YEAR: 1979

DUTY DESCRIPTION
GUARDS AIR FORCE WEAPON SYSTEMS; DEFENDS AIR FORCE INSTALLATIONS; PROTECTS PERSONNEL, EQUIPMENT, AND FACILITIES; AND PERFORMS SPECIAL SECURITY AND LOCAL GROUND DEFENSE FUNCTIONS.

DICTIONARY OF OCCUPATIONAL TITLES

DOD OCCUPATIONAL SUBGROUP 830

SPECIALTY TRAINING STANDARD CENTER LACKLAND TTC

Figure 3.

AFSC 27XXX POPULATION ALL

MAJCOM-SEX SOURCE UAR

MAJCOM	FEMALE	MALE	UNKNOWN
0A= AAC	41	419	0
0J= AFA	1	8	0
0C= ADC	319	2353	0
0D= USAFE	128	1146	0
0E= AFAFC	0	0	0
0F= AFLC	10	70	0
0H= AFSC	43	198	0
0I= ARPC	0	0	0
0J= ATC	209	1054	0
0K= AU	0	4	0
0M= AFRES	0	11	0
0N= HQ USAF	0	55	0
0Q= MAC	214	1226	0
0R= PACAF	62	479	0
0S= SAC	260	1071	0
0T= TAC	367	2173	0
0U= USAFSS	2	19	0
0V= AFCS	427	5639	0
02= AFISC	0	0	0
06= AFAA	0	0	0
09= AFMPC/AFMEA	0	1	0
22= UNK	0	1	0

AFSC 81152

2ND-AFSC SOURCE UAR

ORDER	2AFSC	FREQ
MOST FREQ.		4678
2ND	811X2	434
3RD	811X0	222
4TH	702X0	31
5TH	471X1	8
6TH	231X1	7
7TH	431X2	7
8TH	645XC	7
9TH	645X1	6
10TH	902X0	6
OTHER	OTHER	98

Figure 4.

7014 3PC021 CODAP ANLYS SECURITY POLICE 811X0

DUTY AND TASK PERFORMANCE DATA FOR SELECTED GROUP, BASED ON BACKGROUND INFORMATION LISTED BELOW, FROM THE 811XX CAREER FIELD. TASKS ARE LISTED IN DESCENDING ORDER OF PERCENT TIME SPENT BY ALL GROUP MEMBERS.

UTASK JOB DESCRIPTION	CASES	TASKS	DUTIES	MBRS.
	4508	540	15	1835

Q-TSK	DUTY/TASK TITLE	PERCENT OF MEMBERS PERFORMING	AVERAGE PERCENT TIME SPENT BY MEMBERS PERFORMING	AVERAGE PERCENT TIME SPENT BY ALL MEMBERS	CUMULATIVE SUM OF AVERAGE PERCENT TIME SPENT BY ALL MEMBERS	TASK SEQ NO
E 201	CONTROL ENTRY TO PRIORITY RESOURCES	63.92	4.10	2.62	2.62	
F 250	PERFORM SENTRY DUTY IN RESTRICTED OR CONTROLLED AREAS	55.69	4.64	2.59	5.21	
F 263	RESPOND TO DURESS OR ALARM ACTIVATIONS	68.50	3.44	2.34	7.56	
J 377	PERFORM SECURITY RESPONSE TEAM DUTIES	57.38	3.99	2.29	9.86	
J 360	CONTROL ENTRY INTO OR WITHIN RESTRICTED AREAS OTHER THAN MISSILE SECURITY AREAS	53.84	4.23	2.28	12.13	5
J 367	PERFORM ALARM RESPONSE TEAM DUTIES	53.51	3.87	2.07	14.21	
J 369	PERFORM CLOSE-IN GUARD DUTIES	49.84	4.08	2.04	16.24	
J 368	PERFORM BOUNDARY GUARD DUTIES	49.37	4.03	1.99	18.23	
F 243	OPERATE VEHICLE RADIO OR PUBLIC ADDRESS SYSTEMS	54.27	3.45	1.87	20.10	
O 526	CLEAN WEAPON MECHANISMS OR PARTS	56.95	3.27	1.86	21.97	10
F 192	CLOSE OR SECURE ENTRY POINTS	56.51	3.22	1.82	23.79	
O 540	STORE OR TURN IN WEAPONS	47.03	3.80	1.79	25.58	
O 529	FIRE WEAPONS TO MAINTAIN PROFICIENCY	59.02	3.03	1.79	27.36	
O 537	PERFORM FUNCTIONAL CHECK ON WEAPONS	46.43	3.38	1.57	28.93	
O 533	ISSUE OR RECEIVE WEAPONS OR EQUIPMENT	40.62	3.74	1.53	30.46	15
F 251	PERFORM TACTICAL DEPLOYMENT FROM VEHICLES	48.94	3.09	1.51	31.97	
F 219	GUARD CLASSIFIED EQUIPMENT OR MATERIALS	41.96	3.59	1.51	33.48	
E 262	RESPOND TO ALERTING ORDERS	46.16	3.05	1.41	34.88	
O 535	LUBRICATE WEAPON MECHANISMS OR PARTS	46.98	2.95	1.39	36.27	
O 532	INSPECT OPERATIONAL CONDITION OF WEAPONS	42.13	3.17	1.33	37.60	20
O 538	RECEIVE, STORE, OR RETURN GOVERNMENT OWNED WEAPONS	37.49	3.56	1.33	38.94	
E 194	CONDUCT BUILDING SECURITY CHECKS	47.74	2.78	1.33	40.26	
J 374	PERFORM MOBILE RESERVE FIRE TEAM DUTIES	37.44	3.53	1.32	41.58	
O 523	ACCOUNT FOR AMMUNITION	37.49	3.47	1.30	42.88	
F 199	CONDUCT SEARCHES OF PERSONS OTHER THAN WITH DETECTOR DOGS	52.75	2.26	1.19	44.08	25
E 233	MAINTAIN TELEPHONE OR RADIO COMMUNICATION WITH CENTRAL OR WING SECURITY CONTROL	32.59	3.43	1.12	45.19	
E 197	CONDUCT INSPECTIONS OF POST-ASSOCIATED EQUIPMENT	34.50	3.23	1.11	46.31	
F 200	CONDUCT SEARCHES OF VEHICLES OTHER THAN WITH DETECTOR DOGS	49.10	2.22	1.09	47.40	
J 378	PERFORM WEAPONS CONVOY DUTIES	38.91	2.58	1.00	48.40	
F 190	BRIEF RELIEF SECURITY POLICE ON SPECIAL ACTIVITIES SCHEDULED TO OCCUR WITHIN ASSIGNED AREAS OF RESPONSIBILITY	34.50	2.86	.99	49.39	30
F 242	OPERATE TELEPHONE COMMUNICATIONS EQUIPMENT	33.68	2.91	.98	50.37	
F 215	EVALUATE SITUATION AT ALARM ACTIVATIONS	36.19	2.60	.94	51.31	
J 361	CONTROL ENTRY INTO STORAGE STRUCTURES	29.37	3.09	.91	52.22	
O 525	APPLY PRESERVATIVES TO WEAPONS	28.28	2.80	.79	53.01	
F 271	TEST ALARM SYSTEMS	29.43	2.66	.78	53.79	35

Figure 5.

NUMBER OF TASKS FOR TOTAL SAMPLE

111X0 (6898)

INTERVAL		SPCC01
1.0000	25.9999	1.4
26.0000	50.9999	.9
51.0000	75.9999	.9
76.0000	100.9999	2.5
101.0000	125.9999	8.4
126.0000	150.9999	24.2
151.0000	175.9999	24.4
176.0000	200.9999	17.6
201.0000	225.9999	3.4
226.0000	250.9999	5.6
251.0000	275.9999	2.7
276.0000	300.9999	2.0
301.0000	325.9999	.0
326.0000	350.9999	.0
351.0000	375.9999	.0
376.0000	400.9999	.0
401.0000	425.9999	.9
426.0000	431.0000	.5
MEAN		167.9458
STD DEVIATION		55.3097

111X0 (7758)

INTERVAL	SPC001
TOTAL COUNTED	444.
OTHER	.0
MEAN	174.8536
STD DEVIATION	55.8522

204X0 (7249)

INTERVAL	SPC002
TOTAL COUNTED	882.
OTHER	.0
MEAN	53.6145
STD DEVIATION	38.9749

251X0 (7274)

INTERVAL		SPC001
1.0000	25.9999	12.2
26.0000	50.9999	14.8
51.0000	75.9999	21.2
76.0000	100.9999	22.7
101.0000	125.9999	10.7
126.0000	150.9999	6.6
151.0000	175.9999	4.1
176.0000	200.9999	2.7
201.0000	225.9999	2.0
226.0000	250.9999	1.3
251.0000	275.9999	.7
276.0000	300.9999	.4
301.0000	325.9999	.1
326.0000	350.9999	.0
351.0000	375.9999	.0
376.0000	400.9999	.0
401.0000	425.9999	.0
426.0000	450.9999	.0
451.0000	475.9999	.0
476.0000	500.9999	.0
501.0000	525.9999	.0
526.0000	550.9999	.0
551.0000	575.9999	.0
576.0000	600.0000	.0
TOTAL COUNTED		1900.
OTHER		.0
MEAN		84.3321
STD DEVIATION		53.6695

271X1 (7029)

INTERVAL		SPC002
1.0000	50.9999	30.5
51.0000	100.9999	37.1
101.0000	150.9999	23.1
151.0000	200.9999	6.3
201.0000	250.9999	2.3
251.0000	300.9999	.4
301.0000	350.9999	.2
351.0000	400.9999	.0
401.0000	450.9999	.0
451.0000	500.9999	.0
501.0000	548.0000	.2
TOTAL COUNTED		568.
OTHER		.0
MEAN		81.4595
STD DEVIATION		55.8582

396X0 (7012)

INTERVAL	SPC001
MEAN	103.8906
STD DEVIATION	64.8637

541X1 (6813)

INTERVAL		SPL101
1.0000	50.9999	15.1
51.0000	100.9999	30.8
101.0000	150.9999	17.8
151.0000	200.9999	11.4
201.0000	250.9999	11.4
251.0000	300.9999	8.1
301.0000	350.9999	3.2
351.0000	400.9999	1.6
401.0000	450.9999	.5
451.0000	500.9999	.0
501.0000	550.0000	.0
TOTAL COUNTED		185.
MEAN		134.0432
STD DEVIATION		90.9958

Predicting Trainability of M1 (Abrams) Tank Crewmen

Charlotte H. Campbell
Human Resources Research Organization

Barbara A. Black
U.S. Army Research Institute

The introduction of the M1 (Abrams) tank, with its technologically sophisticated weapon system and correspondingly complex maintenance demands, has caused the U.S. Army to take a hard look at the selection process for M1 training. The current Armor selection standard is a score of 85 or higher on the Combat Operations (CO) scale of the Armed Services Vocational Aptitude Battery (ASVAB); the CO scale is composed of scores on four subtests: Automotive/Shop Information, Coding Speed, Mechanical Comprehension, and Arithmetic Reasoning. An analysis of differences between the M60A1 and M1 tanks (Black & Kraemer, 1981) concluded that while the M1 tank is easier to fire when fully operational and no more difficult to fire when not fully functional than the M60A1, the tasks required for overall operation are more lengthy and critical for the M1. The concern was that more technical skills, or even higher overall intelligence levels, might be needed for M1 crewmen.

The purpose of the research on which this report is based (Campbell & Black, 1982) was to develop and test two types of predictors of the performance of M1 trainees. The first type was to be based on the ASVAB subtests, already being obtained for all soldiers. The second type was based on an approach known as job sample testing. In this context, job sample testing is a means of measuring job or training aptitude by testing part task performance on critical portions of the job.

Data were collected on soldiers in the first two One Station Unit Training (OSUT) classes for the M1, a total of 146 soldiers. Their scores on a special administration of the ASVAB were obtained: ten subtest scores; CO score; the General Technical (GT) score, composed of Arithmetic Reasoning, Word Knowledge, and Paragraph Comprehension; and the Armed Forces Qualification Test (AFQT) score, which is used to determine a soldier's mental category. They were also tested on five job sample tests developed by the Army Research Institute: gunner tracking, target acquisition, operation of the fire control computer, use of the technical manual (TM), and round sensing. Time and accuracy measures were obtained for the tracking, target acquisition, and computer tests, and accuracy for the TM and round sensing tests. The criterion measures included tests results on the Graduate Armor Tests (GATE tests), for tests that covered M1 tasks; target hits during firing of Tank Table VII; and rankings of soldiers by the drill sergeants and tank commanders of the training brigade. Descriptive statistics are at Table 1.

Table 1
Descriptive Statistics for OSUT Criteria

	OSUT I			OSUT II		
Criterion	N	Mean	Standard Deviation	N	Mean	Standard Deviation
GATE Scores	88	87.6%	8.4	58	90.4%	10.5
Firing Hits	82	65.7%	26.3	58	79.6%	17.6
Rankings	88	25.21	9.49	58	25.54	9.26

The first phase of the analysis consisted of a search for ASVAB subtests that could provide better predictions of training success than CO. A comparison of the two OSUT classes on CO, GT, and AFQT scores revealed no differences between the two classes. For OSUT I, CO was the best predictor of GATE scores and rankings, but for OSUT II, AFQT was best (see Table 2). For the two OSUT classes combined, CO was highest. Therefore, CO was designated as the standard to be matched or beaten. (None of the three ASVAB scale scores was significantly correlated with firing hits.)

Table 2
Correlations Between ASVAB Scale Scores and OSUT Criteria

	OSUT I			OSUT II			Combined		
Criterion	CO	GT	AFQT	CO	GT	AFQT	CO	GT	AFQT
GATE Scores	.411**	.237*	.230*	.278*	.231*	.325*	.330**	.206*	.204*
Firing Hits	.066	.129	.181	.143	.128	.077	.053	.076	.106
Rankings	.391**	.296**	.337**	.256	.120	.283*	.338**	.223**	.304**

*p < .05.
**p < .01.

To determine whether any ASVAB subtests could combine to predict training success, multiple regressions were calculated on each of the three criteria, separately for the two OSUT. The separate regressions were performed to provide two sets of predictors so that a double crossvalidation could be carried out, applying the predictors derived from the data of one OSUT to the data of the other. Results are displayed in Table 3. For GATE scores, one analysis selected Auto/Shop Information and Coding Speed, the other selected only Auto/Shop Information. For firing hits, one analysis selected Numerical Operations, the other found no predictors among the subtests. For rankings, one analysis selected Mathematics Knowledge and Paragraph Comprehension, the other selected Numerical Operations and Electronics Information.

Table 3
Results of Regressions of ASVAB Subtests on OSUT Criteria

			Correlations[a]	
Criterion	OSUT	Predictors Selected	OSUT I	OSUT II
GATE Scores	I	Auto/Shop Info. + Coding Speed	.425**	.298* (X)
	II	Auto/Shop Info.	.388** (X)	.358**
Firing Hits	I	Numerical Operations	.278*	-.055 (X)
	II	(no predictors)	-	-
Rankings	I	Math Knowl. + Para. Comp.	.422**	.204 (X)
	II	Num. Ops. + Elec. Info.	.301** (X)	.502**

[a]Correlation with unit weighted predictors. (X) indicates crossvalidation coefficient.
*p < .05.
**p < .01.

The crossvalidations were performed using unit weighted composites of predictors, in which a weight of one is applied to the standardized score of each predictor, and the sum is then correlated with the criterion. Both predictors of GATE scores--Auto/Shop Information and Auto/Shop Information

plus Coding Speed—crossvalidated (i.e., had statistically significant correlations between the unit weighted composite and the criterion), as did Numerical Operations plus Electronics Information as a predictor of rankings.

These four subtests were then combined into a single unit weighted composite, entitled CO-M1. It is highly correlated with CO in both OSUT. When correlated with GATE scores and rankings, it has lower correlations than CO in OSUT I, and higher in OSUT II and in the combined group (see Table 4). The differences between correlations with CO and correlations with CO-M1 are not statistically significant; in the combined group of all soldiers, the difference in the squared correlation, or percent of variance accounted for, is about 2% for GATE scores and 6% for rankings, in favor of CO-M1.

Table 4
Correlations Between CO-M1
and OSUT Criteria

Criterion	OSUT I	OSUT II	Combined
GATE Scores	.390**	.370**	.360**
Firing Hits	.104	-.035	.038
Rankings	.379**	.506**	.421**

*$p < .05$.
**$p < .01$.

Thus there is some indication that CO-M1 may effect a modest improvement over CO in predicting M1 training success. At the same time, there is no evidence in these data that CO is not itself an effective predictor, except that it is not correlated with firing hits.

However, the CO-M1 composite has intuitive appeal for future use. As equipment, manuals, and job aids become more sophisticated, they take over many of the thinking processes formerly required of soldiers, particularly in algebraic manipulations. The soldier no longer uses formulas. He enters a table with certain parameters and finds the necessary solution. Or he enters the parameters into a fire control computer, and the answer is applied to his firing as a correction without him ever knowing it. In some cases he does not enter the input data; many inputs (e.g., crosswind, cant) are sensed automatically. Basic arithmetic, as measured by NO, may be all he needs. Furthermore, the increased sophistication of the M1 tank has relied on vast amounts of electronics equipment. A person familiar with electronics concepts, who does well on EI, may also be the person who quickly becomes comfortable with and proficient on his M1 tank.

But further research relating success in M1 OSUT to CO-M1 is needed before a recommendation to change the selection criterion is justified. This line of research should also be extended to other MOS in Armor (e.g., for Scout and M60 tank crewman training), because: (a) assignment of Armor soldiers trained on one Armor system or for one crew position to another system or position within Armor should not be further complicated by different aptitudes required in different Armor MOS; (b) a different selection criterion only for M1 OSUT would be cumbersome to implement; and (c) technological advances have also been made on other Armor systems such that CO-M1 may be an improvement over CO as an Armor training selector in general.

The second set of analyses focused on the usefulness of the job sample test variables to augment the prediction from CO, or from CO-M1. Again, multiple regressions were used, separately for the two OSUT. This time, however, the analysis forced either CO or CO-M1 to enter the equation first; the relevant job sample test variables were then considered for possible contributions to the regressions. In this way, the job sample tests acted on only that portion of variance in a criterion that was not already explained by CO or CO-M1. In terms of utility, it addresses the predictive power of job sample test variables, given that soldiers are already screened on the basis of the ASVAB. Crossvalidation again consisted of correlations with unit weighted composites of selected predictors.

For the regression on GATE scores, only the computer and TM job sample test variables were considered for inclusion; the tracking, target acquisition and round sensing job sample test variables were added to the regression on firing hits. Test variables and criteria were matched in this way on a rational basis--there is no reason to expect the perceptual-motor job sample tests to reliably predict performance on GATE tests, which cover primarily procedural tasks that are not time constrained, nor is there reason to expect the cognitive job sample tests to predict highly skilled gunnery performance. Because drill sergeant and tank commander rankings were likely based on their knowledge of both GATE scores and firing performance, all job sample test variables were considered for inclusion in the prediction of rankings. Results are displayed in Table 5.

Table 5
Results of Regressions of Job Sample Test Variables on OSUT Criteria

Criteria	OSUT	Predictors Selected[a]	Correlations[b] OSUT I	Correlations[b] OSUT II
GATE Scores	I	CO - Comp. Accuracy	.368**	.368** (X)
Firing Hits	II	CO + Round Sensing	.113 (X)	.347**
	II	CO-M1 + Round Sens.	.143 (X)	.228
Rankings	I	CO - Comp. Accuracy	.442**	.233 (X)
	II	CO - Target Acq.Time	.425** (X)	.377**
	II	CO-M1 - Comp. Acc.	.406** (X)	.311**

[a]CO or CO-M1 entered first. If no job sample test variables were added to CO or CO-M1, the predictor equation is not listed.

[b]Correlation with unit weighted predictors. (X) indicates crossvalidation coefficient.

*$p < .05$.
**$p < .01$.

When CO was entered first, the prediction of GATE scores was improved by the addition of computer accuracy (with a negative regression weight) in one regression--this finding was crossvalidated--but not augmented by any variables in the other regression. When CO-M1 was entered first, no other variables entered the prediction for either OSUT. For firing hits, both regressions in one OSUT added round sensing accuracy to the prediction, but these failed to crossvalidate. In the other OSUT, neither CO nor CO-M1 drew in any

variables, but neither alone was a statistically reliable predictor of hits. In the prediction of rankings, CO was augmented by computer accuracy in one regression and by target acquisition time in the other; only target acquisition time was crossvalidated. With CO-M1, one regression added no variables, while the other added computer accuracy, which was negatively weighted; this result did crossvalidate.

Although none of the job sample tests was able to dramatically improve the predictions from CO or CO-M1 alone, there are some consistent relationships that warrant further examination. The first is the association of computer accuracy with both GATE scores and rankings, always with a negative weight, after CO or CO-M1 has entered the prediction. Either computer accuracy has something in common with CO and CO-M1 that they do not share with the criteria, or computer accuracy is truly negatively related to something in the criteria that is not predicted by CO or CO-M1. What do CO, CO-M1, GATE scores, and rankings have in common? CO and CO-M1 obviously share the subtests Auto/Shop Information and Coding Speed; GATE scores and rankings probably share GATE test outcome. Predictors and criteria share, at a minimum, the knowledge of tools and attention to detail that the ASVAB subtests measure. But CO and CO-M1 are scored very reliably and consistently--an answer is either right or wrong. They also depend on reading skills. GATE tests and rankings, on the other hand, were both somewhat subjective--an obvious point in the case of rankings but also true for GATE tests where scorers give inadvertent cues to correct and incorrect performance. The computer job sample test was also rigidly scored, and demanded a high level of reading ability. Thus computer accuracy is probably only included in the prediction of GATE scores and rankings because it does not predict them, but does make the predictions from CO and CO-M1 more reliable. Its usefulness as a selection tool, if this is the case, would be somewhat limited.

A second consistency is observed in the prediction of rankings from CO and target acquisition time. Even without CO, target acquisition time is correlated highest of all the job sample tests with rankings in both OSUT. This time measure does not necessarily represent accuracy, but more of a quick decision characteristic. It would not be surprising if drill sergeants and tank commanders gave high marks based on their perception of decisive thinking.

Thus, there are indications that the approach is sound, although the desired point-to-point relationship between the job samples and actual performance was not achieved here. Somewhat mixed success has been experienced in using such tests to predict job performance (Eaton, 1978; Eaton, Johnson, & Black, 1980). Additionally, the relationship between trainability and job performance has not been fully explored, and not at all for M1 crewmen. Follow-up of these soldiers after they are assigned to units would provide the opportunity to examine the relationship between job performance, trainability, and job sample testing.

Weaknesses in the present research should be mentioned so that results may be interpreted accordingly, and future work may be better planned. A significant and unavoidable problem concerns the nature of the criteria. Hypotheses concerning the prediction of soldiers' ability to operate the fire control computer could not be tested because a definitive criterion measure of that behavior could not be derived from GATE tests. Criteria against

which to measure the predictive power of the TM job sample test were not available; GATE tests that did require soldiers to use the TM in fact required only that he read aloud given paragraphs in response to scorer questions. Main gun firing data, which were to serve as criteria for the three psychomotor job sample tests, were contaminated (from the researcher's point of view) by admirable (from the trainer's perspective) coaching and assistance from the TC, as well as the simple fact that range conditions did not provide for moving targets and the firing exercise required no tracking, round sensing, or target acquisition. It was, in fact, training and not a test. As such, it provided data that are likely neither valid nor reliable.

If these criteria are measures of what is meant by "success in training," then the conclusion is clear: use either CO or CO-M1 as the selector. These ASVAB composites were both correlated with GATE scores in both OSUT. But until training criteria can be more reliably measured, job sample test results will be of little use in predicting trainability. The fact that the job sample variables did predict some of the variance in the criteria that was not explained by CO or CO-M1 indicates that research on job samples in the Army should not be considered complete.

References

Black, B. A. ASVAB Aptitude Area Score, CO, As A Predictor of Tank Crewmember Performance (Working Paper 80-9). Fort Knox, Kentucky: U.S. Army Research Institute for the Behavioral and Social Sciences, 1980.

Black, B. A., & Kraemer, R. E. XM1 Gunnery Training and Aptitude Requirements Analyses (Research Product 81-5). Fort Knox, Kentucky: U.S. Army Research Institute for the Behavioral and Social Sciences, 1981.

Campbell, C. H., & Black, B. A. Predicting Trainability of M1 Crewmen (Final Report FR-MTRD(KY)-82-7). Fort Knox, Kentucky: Human Resources Research Organization, 1982.

Eaton, N. K. Predicting Tank Gunnery Performance (Research Memorandum 78-6). Fort Knox, Kentucky: U. S. Army Research Institute for the Behavioral and Social Sciences, 1978.

Eaton, N. K., Johnson, J., & Black, B. A. Job Samples as Tank Gunnery Performance Predictors (Technical Report No. 473). Fort Knox, Kentucky: U. S. Army Research Institute for the Behavioral and Social Sciences, 1980.

ABSTRACT

<u>Installation and Testing of CODAP 80 at NODAC</u>

David W. Campbell
John J. Pass, PhD

Navy Occupational Development and Analysis Center

CODAP 80, as received by NODAC from Texas A&M University, is in an IBM OS/JCL "ready to go" state, and as such, can be easily installed on an IBM OS system. To use CODAP 80 on a non-OS/JCL system, the machine interface must be converted. Machine interface is a definition to the computer of the various files being used and of the programs to be executed. OS/JCL is a means of doing that for IBM operating systems (OS); and "CMS EXEC" is a means for VM/370 (virtual machine). This paper described one alternative set of steps to follow to install the OS/JCL version of CODAP 80 onto an IBM 4331 CMS (VM/370) system. The paper also described the steps taken to test CODAP 80, including creating CODAP 80 output to match against standard CODAP output. Displays of actual codes were used. Comments on the flexibility and the utility of CODAP 80 were also presented.

United States Army Advanced Medic (91B30) Training: An Iterative Decision Method Application

CPT Terry D. Carroll, M.A., M.Ed., Ph.D and
Kenn Finstuen, M.S., M.Ed., Ph.D.
Academy of Health Sciences, United States Army
Fort Sam Houston, Texas 78234

Abstract

This paper describes the first implementation of the Iterative Decision Method (IDM) for the selection of training tasks in the 91B30 Advanced Medical Specialist Course, US Army Academy of Health Sciences. The purpose of this research was to determine the feasibility of conducting front-end-analysis of medical training requirements with the IDM. Five expert judges were employed to select or nonselect 209 tasks, grouped into 13 modules, ranging from 3-58 medical tasks. In the first iteration, judges made independent selection decisions (J1). Task judgments were analyzed for goodness-of-fit (*R*) and inter-rater reliability (r_{kk}). Next, judges met and reviewed the results. Discussion was directed to disagreed upon tasks. Revised group judgments (J2) followed, with consequent increases in *R* and r_{55}. For the largest module, Medical/Surgical Procedures, findings indicated J1-J2 increases of .55 to .93 for *R*, and .38 to .96 for r_{55}. Finally, tasks were prioritized within modules based on 3-point task training ratings.

Background

The Academy of Health Sciences (AHS), Fort Sam Houston, TX, has the responsibility for the development and implementation of training for over 30 enlisted medical military occupational specialties (MOS). Within the Academy's organizational framework, the Directorate of Training Development (DTD) holds primary purview for the delineation of training requirements for jobs and tasks within each medical MOS, and, in conjunction with the Directorate of Combat Development and Health Care Studies (DCDHCS), has the responsibility for revising training programs to meet emerging combat medical needs. The largest and most significant MOS which the Academy trains is the 91B Medical Specialist, with over 15,000 active and 22,000 reserve component positions authorized (7th largest MOS in the US Army). Prior training for this MOS consisted of a single Advanced Individual Training (AIT) phase ranging from 6 to 10 weeks. The possibility existed that a 91B medic could complete a 30-year career with only AIT and no additional mid-career MOS training. For any technical field, and in particular medical jobs, the resultant training deficiency is obvious. Further, analyses conducted by DCDHCS were conclusive in the identification of the need for combat medics to acquire new and sophisticated trauma skills for the treatment of casualties on middle to high intensity battlefields.

To remedy these problems, The Surgeon General of the Army, in February 1981, directed the Academy to develop a new Advanced Medical Specialist Course. An implementation date of April 1983 was targeted for the new 91B30 program.

The central problems confronting the developers of the 91B30 course consisted of the identification of job performance criteria, and the selection of tasks to be trained. Utilizing the Instructional Systems Development (ISD) technology (TRADOC, 1975), a number of task lists were prepared by various teaching elements within the Academy, viz., Medicine and Surgery, Physicians Assistant, and Special Forces Aidman. These lists were compiled by DTD and an initial Critical Task Selection Board (CTSB) was convened. Meeting twice in September 1981, the board selected 220 of 443 medical tasks for training. The board consisted of 20 Army Medical Department (AMEDD) personnel, 10 officers (0-3 to 0-6) and 10 enlisted (E-6 to E-9).

A number of problems were encountered with the CTSB configuration, but the most significant areas were: a) the board spent inordinate amounts of time discussing items on which they agreed; b) rank and branch of service, rather than experience and expertise often influenced decision making; c) individual

participation was limited due to the size of the group; and d) semantic problems, particularly across professional lines, occurred frequently. Problems not withstanding, the initial 91B30 task list was reviewed and sanctioned by the AHS Commandant, 29 September 1981. While the task list contained numerous critical life saving duties, many Army medical professionals felt that the list was incomplete and additional tasks were requested to be added to the list by the Office of The Surgeon General (OTSG) and OTSG consultants.

The list underwent continued refinement during a Site Device Selection Board (SDSB), required by the ISD process, which was held in February, 1982. The SDSB recommended further semantic changes to task titles and added another 16 tasks to the list. The lack of an acceptable quantitative method for task selection and prioritization made it increasingly difficult to stabilize the task list. As a result, the list was subjected to many additional alterations and modifications. In short, closure was needed on the task list to eliminate the recurring amendment process before a final list could be sanctioned by OTSG. To meet this need the Iterative Decision Method (IDM) was developed (Finstuen, 1982; Note 1) and plans were made to test the technology.

Method

Participants

The first major step in implementing the IDM involved the procurement of five expert medical judges to serve in the process. To insure balanced results, OTSG input, Reserve component participation, and Academy Directorate representatives were required. Recommendations from the OTSG consultants on emergency medicine and emergency nursing were requested and an Emergency Medical Service (EMS) physician and Emergency Room (ER) nurse were cited, by name, to participate on the board. Through the National Guard Liaison Office, AHS, an approved Reserve Component 91B incumbent was secured. In addition, the Academy provided two senior NCOs, from the Directorates of Training and Training Development. The five board members constituted the 91B30 Critical Task Relook Board.

Materials and Procedure

The 91B30 task list consisted of 209 tasks, and was divided into 13 duty modules. Modules ranged from 3 to 59 tasks. For the purposes of this paper, the largest and most significant segment, Medical and Surgical Procedures, will be the only detailed module presented. Other modules included topics such as field sanitation, preventive medicine, and combat psychiatry. Overall results also will be included. A detailed technical report covering all aspects of the project is in progress and will be available from DTD at a later date. Table 1 presents examples of some of the medical and surgical procedural tasks.

Table 1
Examples of Medical and Surgical Procedure Tasks in the 91B30 Advanced Medic Course

No	Task Statement
6.	Perform Suturing Techniques
7.	Control Hemorrhage by Ligation of Vessels
8.	Control Hemorrhage by Clamping of Vessels
16.	Don Sterile Gown and Gloves
19.	Identify and Manage Multiple System Trauma
20.	Identify and Manage High Velocity Missile Wounds
28.	Perform EOA, Nasal, and Endotracheal Tube Insertion
29.	Perform Cricothyroidotomy
31.	Perform Chest Decompression
34.	Operate and Maintain Suction Equipment
41.	Perform Urinary Catheterization
57.	Apply First Aid to a Patient With Anaphylactic Shock

A briefing was prepared and presented to each of the participants outlining their mission, and the basic technology of the method. Judges were encouraged to participate in the process regardless of their position on any issue viz-a-viz other judges.

The IDM is a highly structured group judgment model, designed to maximize the effectiveness and efficiency of decision making, for a panel of 5 or 7 experts. The technology draws from several decision making techniques (i.e., Nominal Group Technique and Delphi Processes, Delbecq, Van de Ven, & Gustafson, 1975) and is based upon the research findings of over 70 small group interaction and productivity studies

(Finstuen, 1982). The productivity of the IDM process rested on two critical tenets. First, to maximize effectiveness, independent judgment (J1) results, from a nominal group were used as feedback for making the revised group judgments (J2) under a "pooling-of-abilities" model. Numerous research investigations have shown that discussion and revision of group judgments increases the accuracy of the decisions (Huber & Delbecq, 1972; Shaw, 1971, Steiner, 1972; Thorndike, 1938) and are more motivating and satisfying to participants than purely nominal group judgments (Hackman & Morris, 1975; Hare, 1962; Shiftlett, 1972).

Multiple linear regression equations (Ward & Jennings, 1973) were used to express decisions of the nominal group as a function of dichotomously coded task and rater variables. Group equations for each duty module took the following form:

$$Y = w_1T^{(1)} + w_2T^{(2)} + \ldots + w_nT^{(n)} + w_{(n+1)}R^{(1)} + \ldots + w_{(n+k)}R^{(k)} + c,$$

where Y was a criterion vector of decision scores (length equals $\underline{k}$ raters times $\underline{n}$ tasks), $T^{(i)}$, i = 1 to $\underline{n}$, was a task predictor variable coded 1 if decisions were observed on task i, 0 otherwise; $R^{(j)}$, j = 1 to $\underline{k}$, was a rater predictor variable coded 1 if decisions were associated with rater j, 0 otherwise; w_1 through $w_{(n+k)}$ were the raw least squares regression weights associated with each predictor, and c was a regression constant. Selection criteria consisted of binary decision scores (Lunney, 1970) and were coded 1 if a task was selected for training, 0 if nonselected. Multiple correlation coefficients, $\underline{R}$'s, were used as indicators of the goodness-of-fit for the group prediction equations.

Second, to increase efficiency, discussion was directed to disagreements which merited attention, and not to tasks which the experts had already agreed upon for either selection or nonselection. The gross level of group agreement for duty modules was measured by the inter-rater reliability coefficient $\underline{r}_{kk}$ (Guilford & Fruchter, 1973). Specific task and rater disagreements were identified by examining the squared residual contributions of task and rater variables to the total squared residuals associated with the group equation. With this form of decision making there were no correct or incorrect expert opinions. The objective of the process was to have the group arrive at an acceptable level of agreement in regard to the tasks selected for training; it was not necessary that 100% consensus be obtained. After tasks were selected for training, they were prioritized and categorized through the use of an anchored 3-point combat criticality rating scale (3 = combat critical--tasks crucial to survival in combat; 2 = mission essential--tasks necessary to support the stated mission of peacetime AMEDD organizations; and 1 = other essential--tasks that contributed to the performance of combat critical or mission essential tasks, but did not, by themselves, affect mission attainment).

Clearly, this technology remedied several of the key problems experienced with the CTSB, but most noteworthy was the assurance that all expert judges contributed their expertise individually and as group members, and that the selection decisions were made in an effective and efficient manner. It was anticipated that the technology would provide the needed closure through the stabilization and prioritization of the task list, based upon judgments secured from the medical expert judges.

Data collection began 23 April 1982, by securing independent task selection judgments (J1) from the Academy members and the Reserve Component representative. On 27 April 1982, an AHS team traveled to Darnell Army

Hospital, Fort Hood, TX, to gather data from the EMS physician and ER nurse. The group component of the IDM (J2) was secured 6-7 May 1982 at Fort Sam Houston. DTD sponsored the assembly of all of the judges, and after a review of the J1 findings and procedural briefings, J2 judgments were rendered.

Several actions taken at the convention of the board were of particular assistance to the members. First, to provide a frame of reference for decision making, DCDHCS presented a briefing on the scenario of the modern battlefield and the equipment the 91B30 would have to use. Second, results from an initial front-end-analysis (FEA) of the task list items were made available by several 91B30 subject matter experts. Third, representatives from Collective Training Division, DTD, and DCDHCS were on hand to answer technical questions relating to the needs and requirements of the Army in general. Finally, the project officer served as facilitator to insure smooth procedural operation.

Results

Collectively the board had 70 years of active duty Army medical experience, of which 39 years had been served in Table of Organization and Equipment (TOE) field units. In addition, two enlisted members of the board had combat experience and had collectively served a total of 39 months in Viet Nam. On the average, board members were 35 years old, and had an average of 16 years of formal education.

Selection of Tasks for Training

A summary of the overall J1-J2 selection results and prioritization results is presented at Table 2, together with specific results obtained for the Medical and Surgical Procedures Duty module. As shown, some 97% (100 x .97) of the 290 medical and surgical J1 task decisions were voted as "select". Goodness-of-fit for the group equation ($\underline{R}$ = .55) was modest and the low reliability (.38) for this module indicated that group discussion was required. Figure 1 presents the standardized display, which experts used to interpret disagreements for the medical/surgical duty.

Table 2
Summary of Independent (J1) and Revised Group (J2)
91B30 Task Selection Judgments[a] and Combat Criticality Ratings[b]

Decisions	# Tasks	$\underline{k} \times \underline{n}$ Judgments[c]	Mean	S.D.	$\underline{R}$[d]	$\underline{r}_{55}$
J1 Independent Judgments						
Medical Surgical	58	290	.97	.16	.55	.38
Overall	209	1,045	.93	.26	.56	.40
J2 Revised Group Judgments						
Medical Surgical	58	290	.98	.14	.93	.96
Overall	209	1,045	.97	.16	.63	.61
Combat Criticality Ratings						
Medical Surgical	57	285	2.71	.58	.80	.84
Overall	207	1,035	2.56	.40	.95	.97

[a]Judgments were coded 1 if a task was selected, 0 if nonselected
[b]Prioritization ratings were obtained from 3-point rating scales ranging from 3=combat critical, 2=mission essential, to 1=other essential.
[c]$\underline{k}$=5 experts
[d]Multiple correlations were statistically significant from zero, $\underline{p}$<.01

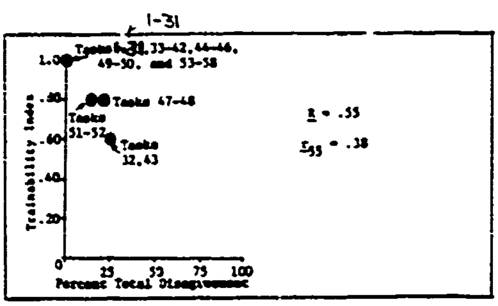


Figure 1. Standard graphic display of J1 independent judgment results for the Medical and Surgical Procedures module.

As shown, task selection averages (trainability indices) ranged from 0 to 1.0 and were plotted vertically. Task information was also plotted horizontally in terms of the amount of disagreement each task exhibited (percent of each task's squared residual sum in relation to the total group equations' sum of squared residuals). Most tasks, clustered in the upper left corner, were selected for training and all raters agreed they should be selected (zero disagreement). However, Tasks 32 (Perform Thoracentesis), 43 (Perform Advanced Cardiac Life Support), 47 and 48 (Pertaining to Pediatrics and Child Abuse), and 51 and 52 (Snake Bite and Antivenom) were disagreed for selection.

After discussion of those particular tasks, the board rendered a revised set of judgments. One task (51) was declared as nonselect by

all members of the board, and four raters decided to select Task 43 while one did not. Because one expert still disagreed on this task, its selection priority resulted in .80. Both goodness-of-fit and inter-rater reliability (r_{kk}) substantively increased for the revised group judgments as a result of the discussion (to .93 and .96 respectively).

This finding indicated that the information exchanged during the revised group judgment phase produced a more carefully considered and agreed upon listing of training tasks, even though 100% consensus was not attained. After the revised group judgments were made, the tasks selected for training (207 out of 209) were rated using a 3-point combat criticality scale (Table 2). Findings for medical and surgical procedures, and for all the modules, indicated that the ratings were stable and reliable. Table 3 presents the results for hypothesis tests of differences among task selection and prioritization averages. These results were used to gauge the effects of task variables in regard to the dependent decision measures, while controlling for the effects due to raters. Full group equation results (R^2_{full}) were tested against results from equations restricted to only rater variables ($R^2_{restricted}$). Significant results were obtained for all comparisons, and as shown, differences among task selection means increased from the J1 to the J2 condition. These findings indicated that raters had differentiated among tasks in terms of selection and combat critical priority, and that the group discussion had indeed enhanced the decision making process for the Medical Surgical module, and overall modules.

Table 3
Summary of Significance Tests for
J1-J2 Selection Decisions and Combat Criticality Ratings

Judgment Condition	k x n Judgments	R^2_{full}	R^2_{rest}	df_1	df_2	F*
Medical Surgical Procedures						
J1 Independent Judgments	290	.301	.020	57	228	1.60
J2 Revised Group Judgments	290	.866	.002	57	228	25.82
Combat Critical Ratings	285	.548	.080	58	224	6.45
Over All Modules						
J1 Independent Judgments	1,046	.311	.023	208	832	1.57
J2 Revised Group Judgments	1,045	.481	.017	208	832	2.56
Combat Critical Judgments	1,035	.995	.047	206	824	32.32

*All F ratios were significant, $p < .01$

Table 4 presents an abbreviated prioritized list of the medical and surgical tasks that were selected for training development. Cut-off points were established to group tasks into three categories as shown. The final overall task list contained 74 combat critical, 109 mission essential, and 24 other essential tasks. Tasks which are identified as combat critical, and certain high priority mission essential tasks, are typically employed as input to soldier's field manuals and serve as a basis for specialty qualification testing. Medical and surgical procedures accounted for 26 of the 74 (35.14%) combat critical tasks. While all 207 selected tasks were grouped throughout the range of possible criticality from .6 to 3.0, finer discriminations would probably be desirable. Future studies would benefit from the use of an expanded 7- or 9-point rating scale or a ranking procedure to determine finer just-noticeable-differences among tasks.

Conclusions

The IDM technology provided the DTD with an effective and efficient method of task selection and prioritization and, in the case of the 91B30, task reaffirmation. Through the combined J1 - J2 decision making process and ratings of selected tasks, over 3,000 expert judgments were directly applied to the task data. The prioritized task list constituted a defensible and comprehensive basis for the identification of training requirements and for the subsequent development of training materials and courseware for the 91B30 Advanced Medical Specialist School.

Yet another significant facet of the technology, of considerable import and utility to trainers, was the ordering of duties and tasks within the list. Given the five judges, each task received a rating from 0 non-select, to 1.0, select, separated by intervals of .2. Thus it was possible to group tasks

Table 4

Selected Medical and Surgical Training Tasks
Arrayed by Criticality Index and Category

Original Task No	Task Description	Criticality Index	Category
6.	Perform Suturing Techniques	3.0	
10.	Apply Surgical Dressings and Drains	3.0	[illegible]
19.	Identify and Manage Multiple System Trauma	3.0	
29.	Perform Cricothyroidectomy	3.0	Critical
31.	Perform Chest Decompression	3.0	
13.	Prepare Surgical Patient, Perform Minor Surgical Scrub	2.8	
15.	Set up and Maintain Sterile Field	2.8	
30.	Operate and Maintain O_2 Equipment	2.6	Mission
49.	Identify and Manage Bites of Snakes and Animals and Stings of Insects	2.4	Essential
16.	Don Sterile Gown and Gloves	2.4	
56.	List Effects of Common Poisons	2.0	
46.	Prehospital Childbirth Procedures	1.8	
48.	Identify and Manage Child Abuse Problems	1.8	Other
58.	State Source of Information for Poisons not Commonly Encountered	1.8	Essential
43.	Perform Advanced Cardiac Life Support*	.8	

*Task 43 was appended to the list and indexed by its selection priority value.

with similar trainability index values, i.e., .8, .6, .4, .2, and utilize the selection values in conjunction with the priority ratings as task discriminators, if time or monetary resources precluded the training of all tasks. The a priori statement that 100% consensus of task selection during J2 was not required, provided the expert judges with an opportunity to express their opinions in a way that could change training priorities without completely deleting or adding the task for training (Task 43), an aspect that the judges felt was most equitable.

While this first implementation of the IDM at the Academy served as a relook for tasks that had already been through two boards, the value and workability of the system was established beyond any doubt. In fact, use of the IDM under these circumstances provided a very rigorous test for the technology since the J1 task list had already been refined from a larger original list of 443 medical tasks, so decisions required a high degree of discrimination on the part of the expert judges.

In conclusion the IDM has enormous application potential in any performance technology based organization, but is particularly germane to military training for several reasons. First, the quantifiable aspects of collective expertise provide multiple benefits, with a clear audit trail and statistical soundness providing proper task list closure, not the least of them. Second, the expert judges involved in the methodology can provide inter-agency input equivalent to several iterations of normal staffing. Third, a clear course of action for review/revision protocols consistent with initial action can be provided through subsequent boards.

Reference Note

1. Finstuen, K. An iterative decision method for selecting medical tasks for training. Paper to be presented at the 24th Annual Conference of the Military Testing Association, San Antonio, TX, November 1982.

References

Delbecq, A., Van de Ven, A., & Gustafson, D. Group techniques for program planning: A guide to nominal and Delphi processes. Glenview, Ill.: Scott, Foresman, and Company, 1975.

Finstuen, K. Prioritization of medical combat deficiencies: Application of the iterative decision method (AHS Monograph 82-1). Ft Sam Houston, Tx: Academy of Health Sciences, Directorate of Training Development, September, 1982.

Guilford, J., & Fruchter, B. Fundamental statistics for psychology and education (5th ed.). New York: McGraw-Hill, 1973.

Hackman, J., & Morris, C. Group tasks, group interaction process and group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol 8). New York: Academic Press, 1975.

Hare, A. Handbook of small group research. New York: The Free Press, 1962.

Huber, G., & Delbecq, A. Guidelines for combining judgments of individual group members in discussion conferences. Academy of Management Journal, 1972, 2, 161-174.

Lunney, G. Using analysis of variance with a dichotomous dependent variable: An empirical study. Journal of Educational Measurement. 1970, 7, 263-269.

Shaw, M. Group dynamics: The psychology of small group behavior. New York: McGraw-Hill, 1971.

Shiflett, S. Group performance as a function of task difficulty and organizational interdependence. Organizational Behavior and Human Performance, 1972, 7, 442-456.

Thorndike, R. The effect of discussion upon the correctness of group decisions, when the factor of majority influence is allowed for. Journal of Social Psychology, 1938, 9, 343-362.

TRADOC Pam. 350-30. Interservice procedures for Instructional Systems Development: Phase I Analyze (Vol I). Ft Benning, Ga.: U.S. Army Combat Arms Training Board and Center for Educational Technology, Florida State University. (Contract Nr. N-61339-73-C-0150), Interservice Committee for Instructional Systems Development (ISD) August, 1975.

-ard, J., Jr., & Jennings, E. Introduction to linear models. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

USAMPS TEST VALIDATION PROCEDURES

Dr. Albert L. Castelli

United States Military Police School

Over the past two years the US Army Military Police School has made a concerted effort to improve their resident courses' testing program. This has resulted in a very viable testing program. Before discussing the validation procedure used I will present an introduction of the directorates involved with the testing program. Then I will show how testing fits into the course validation process. Then a detailed discussion of the test validation process will follow.

The Military Police School has three directorates involved with the test validation process; that is, the Directorate of Training Developments (DTD), the Directorate of Training and Doctrine (DOTD) and the Directorate of Evaluation and Standardization (DOES). The DTD is responsible for the ongoing validation of the testing program. DOTD personnel develop the test items and compile the tests. DOES is responsible for the internal and external validation of the testing and instructional program.

Personnel from each of the directorates are part of the Test Review Committee involved with the test validation process. The DOTD is represented by the course manager and the subject matter experts, the DTD by the POI manager and an education specialist, and the DOES has a representative assigned to each resident course.

Before I go into the specifics of the test validation let's review a transparency which depicts the course validation process and note where test validation fits into this process (see Figure 1). You will note that test strategy is discussed after the needs and job/task analysis. The SME and DTD plan the test strategy followed by the SME developing the test items. The development of instruction and the implementation will be in progress while the validation process is functioning.

There are three steps in the validation process (see Figure 2). You will note that the first step is content validation, the second step is a small group trial and the final step, a large group trial or trials, completes the validation process.

Content validation is conducted immediately after the test items have been written by the SMEs. Then the Test Review Committee is convened to review and evaluate the test items. The items are analyzed for their adequacy in testing the students knowledge of a specific task in the module. (Each critical task or critical steps in a task should be tested). Secondly, the items are analyzed for the level of fidelity (see Figure 3). The highest fidelity

possible must be the goal. However, the available resources and time available have to be considered. Therefore, items are evaluated for both grammatical form and meaning. Any apparent ambiguities are changed to improve clarity. Lastly, the items are checked for doctrinal exactness. The SME is asked to furnish specific references that support the correct answer. Any suggested changes are made before the compiled test is subjected to the next step, the small group trial.

The small group trial is under the direction of the POI manager. He selects a specific number of masters and non-masters to whom to administer the test (our criteria is at least five of each category). Masters, as used in the validation process, refers to individuals who by virtue of training or experience should be capable of passing the test. Non-masters refer to those individuals who by virtue of training and or experience should fail. At least 80% of the masters should pass and 80% of the non-masters should fail. The results of the small group trial are submitted to the education specialist, Test Branch, for analysis. The Test Review Committee is again convened to discuss this analysis. At this meeting more changes and revisions may be made. If any changes are major, these specific changes are subjected to another small group trial. After all necessary adjustments are made and approved by the Test Review Committee the test is ready for a large group trial.

The large group consists of a class that has received instruction on the specific module. The test is then administered to this group. The test is graded by the SME and delivered to the education specialist, Test Branch, for another analysis. The test is item analyzed and each item that falls below an 80% difficulty level is tagged for discussion by the Test Review Committee. The education specialist completes a form listing these items with their corresponding difficulty level and discrimination index. The POI manager distributes a copy to each member of the Test Review Committee and schedules a post test analysis meeting. Each one of the items are discussed for validity of the specific item. If a certain incorrect distractor seems to be popular it will be thoroughly analyzed. If it is selected by those who made the lower scores it is possibly a good distractor. The analysis may result in acceptance of the test items, a slight revision of the item or the distractor, a major revision or writing a completely new test item.

A test is finally considered valid if 80% of the class scores 80% or higher and 80% of the test items receive a difficulty level of 80% or higher. This is commonly referred to as the 80-80 criteria.

Each subsequent large group administration is item analyzed and necessary refinements made.

COURSE VALIDATION PROCESS

NEEDS ANALYSIS	JOB/TASK ANALYSIS	TEST STRATEGY
TEST VALIDATION	DEVELOPMENT OF INSTRUCTION	IMPLEMENTATION
CONTENT S-GROUP		
TEST VALIDATION	INTERNAL VALIDATION	EXTERNAL VALIDATION
L-GROUP PERF TEST		

FIGURE 1

VALIDATION STEPS

1. CONTENT

2. SMALL GROUP TRIAL

3. LARGE GROUP TRIAL

FIGURE 2

FIDELITY LEVEL		TYPES OF MEASUREMENT
LOW FIDELITY	1	ASK FOR OPINIONS
	2	ASK FOR ATTITUDES
	3	MEASURE KNOWLEDGE
	4	MEASURE RELATED BEHAVIOR
	5	MEASURE SIMULATED BEHAVIOR
HIGH FIDELITY	6	MEASURE "REAL LIFE" BEHAVIOR

FIGURE 3

AD P000857

DEVELOPMENT OF THE ARMY RESEARCH INSTITUTE INTEREST SURVEY (ARIIS)

John G. Claudy, Ph.D.
John S. Caylor, Ph.D.

American Institutes for Research
Palo Alto, California 94302

Richard A. Kass, Ph.D.

U. S. Army Research Institute

Introduction

Brief Background

Measures of vocational interests have for many years been recognized as important in predicting success in military training and work, and several different types of such instruments have been employed in military settings (for example, the Army Classification Inventory (ACI) and the Air Force VOICE). As tests of vocational aptitude have become increasingly refined, the potential net contribution of non-cognitive measures to efficient, effective prediction of training and job success has grown larger.

As presently constituted, the Armed Services Vocational Aptitude Battery (ASVAB), used by the Army for Military Occupational Specialty (MOS) assignment decisions, does not contain any non-cognitive vocational scales. However, until recently four such scales were included: Maintenance Scale, Electronics Scale, Attentiveness Scale, and Cround Combat Scale. The U. S. Army Research Institute for the Behavioral and Social Sciences (ARI) supported this project to develop a more comprehensive and differentiating measure of non-cognitive vocational interests, called the Army Research Institute Interest Survey (ARIIS), to assist in classification and assignment decisions.

Requirements for a New Test

To permit more specific measurement of vocational interests for a greater number of job areas, the research reported here was undertaken to develop the ARIIS to meet the following requirements:

1. All items are to be appropriate in content and language to the knowledge and experience of first-tour applicants to the Army.

2. The set of items must cover the full range of vocational interests most applicable to Army jobs - with emphasis on high density Army MOS clusters such as armor, clerical, infantry and mechanical.

3. The interest survey must provide an independent, numeric score on vocational interest for each MOS for which a scoring key is developed.

4. The interest survey must permit the development of additional vocational area/MOS keys, as may be needed, without any change in the survey form itself and without further conceptual analysis of the new MOS.

5. The interest survey must permit the development of separate keys for males and females without any change in the survey form itself.

6. All items must be objective in format, with responses chosen from fixed alternatives on a single, machine-scoreable answer sheet.

7. The test should require no more than one hour to administer.

8. The test should be suitable for computer administration and scoring.

Model Selection and Item Development

Model

The initial activity undertaken by the project staff was a comparison and evaluation of various approaches to measuring vocational interests. Instruments designed to measure vocational interests can be classified according to several different dimensions, each of which represents an issue related to the design and development of the instrument. Among the dimensions are:

How the items in the instrument were selected:
- A. Empirical keying
- B. Homogeneous keying
- C. Logical keying

How many scales a single item contributes to:
- A. One scale
- B. Several scales
- C. All scales

How the items are presented to the examinee:
- A. One item at a time
- B. Pairs of items
- C. Triads of items

What sort of choice the examinee is to make:
- A. Degree of liking of the item
- B. Forced choice between items

Type of scoring weights used:
- A. Zero-one weights
- B. Multiple whole number weights
- C. Fractional weights

Suitable for hand scoring:
- A. Yes
- B. No

Type of scores reported:
- A. Broad interest area profiles
- B. Scores on specific occupations

These dimensions are not mutually exclusive, and there are other dimensions along which interest measurement instruments could be classified, however, once the target population for the instrument has been defined, these are the major issues to be addressed. Within these dimensions several tradeoffs exist. For example, if the instrument is designed for easy hand scoring then usually zero-one scoring weights are used, each item contributes to only one or at most a few scales, and scores are usually reported for broad interest areas rather than for specific occupations. On the other hand, if scores on a number of specific occupations are to be reported, then individual items usually contribute to several scales and computer scoring is used. When scores are reported on specific occupation scales, empirical keying of items is usually employed; but when scores are reported on scales for broad interest areas, homogeneous or logical keying is generally employed. In addition, two of the dimensions are so closely tied together as to in effect represent a single issue. If items are presented one at a time, then the examinee virtually always responds by indicating a degree of liking of the item. If the items are presented in pairs or triads, then the examinee responds by making some sort of forced choice between items.

Two of these issues were settled in the initial requirements for the new survey instrument: scores would be reported for specific occupations and hand scoring would not be required. This in effect also specified that empirical keying would be used and that each item would contribute to several or all of the scales. This left only the issues of single vs. multiple item presentation and the types of scoring weights to be used.

The single item presentation approach is best exemplified by the Strong-Campbell Interest Inventory which presents one item at a time and asks the examinee to respond to each item by marking like, dislike or indifferent. The multiple item presentation, forced choice approach is best represented by the Kuder Occupational Interest Survey, Form DD which presents items in triads and asks the examinee to respond by marking the most liked and the least liked of the items (options). In scoring the Strong-Campbell, each item has a scoring weight of +1, 0, or -1 for each scale. The Kuder, Form DD uses a more complex scoring procedure based on fractional weights. Either approach is equally suitable for computer scoring. Thus the selection of an approach to use for the ARIIS was essentially a choice between single item presentation and forced choice presentation.

After comparative study of the various approaches to conceptualizing and measuring vocational interests, it was decided that the requirements for the Army Research Institue Interest Survey (ARIIS) would best be met by developing the survey form in accordance with the recent work of Frederic Kuder (1977), as exemplified in the Kuder Occupational Interest Survey, Form DD (Kuder & Daimond, 1979). That is, a forced choice, triad format.

Following this model, the basic element in the interest survey form is a brief statement of an activity such as "repair a light socket" or "take care of farm animals." Such statements are presented in sets of three, called triads. For each triad, the respondent chooses the one activity in the triad that he/she would MOST like to do and the one activity in the triad that he/she would LEAST like to do.

The data necessary to develop an empirical key for virtually any job (MOS) can be obtained by administering the interest survey to experienced, satisfied incumbents in the job. Scoring weights are established by assigning to each statement in the triad the proportion of job incumbents who most liked that statement, and similarly, the proportion of job incumbents who marked the statement least liked. An individual's scores on such an interest survey can be obtained by comparing that individual's pattern of responses on the triads with the pattern of responses of the job incumbents. The measure of similarity between the responses of job incumbents in any particular MOS and the examinee's responses is the Clemans' Lambda Coefficient (Clemans, 1958). At this point it is sufficient to note that Lambdas range from -1.00 to +1.00 with positive scores indicating a positive relationship, zero scores indicating an absence of similarity or a random relationship, and negative scores indicating a negative relationship between the respondent's preferences, as indicated by responses to the interest survey triads, and those of the job incumbents.

The major advantages of this model are:

1. Only limited, relative judgments are required when alternatives are presented in triads as opposed to the repetitive judgment of all alternatives on an absolute scale.

2. It is comprehensive and efficiency in testing, in that all alternatives are used on all scoring keys.

3. It is easy to develop empirical keys for any other job or MOS without changing the survey form.

4. It provides a numerical index of occupational interest for each job area which facilitates comparing different jobs for an individual as well as different individuals for a job.

Developing an Item Pool

There were two primary considerations in the initial drafting of items for the ARIIS:

1. Items should incorporate choices or preferences that, on a judgmental/conceptual basis, would most clearly differentiate among satisfied and dissatisfied individuals in the major, high density Army MOS clusters.

2 Items should be appropriate in both content and language to the Army applicant population.

Based on these general considerations, the project staff developed the following rules to guide the writing of triads:

Rules for Writing Triads

1. None of the choices should be the names of occupations or professions.

2. All choices should start with an active verb if possible rather than the verb "to be."

3. All choices should be activities a 17 to 20 year old has probably done or has some realistic understanding of.

4. All options in a triad should be equally socially attractive.

5. The activities contained in the options for any given triad should require about the same amount of time to carry out when they are actually done.

6. Some triads may include the same verb with three different objects or modifiers.

7. Some triads may include the same object or modifier with three different verbs.

8. Options should not by their nature exclude females.

9. Use simple, straightforward words and language.

10. Keep options as short as possible.

11. Across triads, options should cover a wide range of behavior and preferences.

Developmental Tryout

Using these guidelines, project staff members independently drafted and jointly examined an initial pool of triads. After revision, editing, and elimination of duplicate and nearly duplicate triads, the remaining 225 triads were divided into three sets of 75 triads each. Each set, along with instructions and a disclaimer stating that participation was voluntary and that results would not be reported to school officials, constituted one of the three developmental tryout forms of the interest survey.

In April of 1980 these three developmental tryout forms were administered to 16 senior students in high school vocational classes in Seaside, CA. Six students took Tryout Form I, five took Tryout Form II, and five took Tryout Form III. Testing times ranged from 7 to 20 minutes with a median of 15 minutes. The participants reported having no trouble understanding the directions or marking the options. All triads on all forms were properly marked and none were omitted.

Based on this limited tryout with high school students, the 225 triads were reviewed and 88 (39 percent) of the triads were revised. The principal revision was to make highly "most preferred" options less attractive and highly "least preferred" options more attractive. This was done in the interest of increasing the potential discrimation power of the triad. All 225 triads were then combined into a single pool for later use.

Data Collection and Triad Selection

Data Collection Booklet

In March of 1981, the project director spent six days in the Republic of Panama where he worked with a team from ARI collecting data from troops of the 193rd Infantry Brigade. In preparation for this field work, the project staff developed a 19 page data collection booklet entitled the "U. S. Army Experimental Classification Inventory." In addition to the 225 interest triads

developed as a part of this contract, the data collection booklet contained two sections of questions provided by ARI. The first of these was called the Job Performance Self-Report and contained 12 items related to the soldier's reasons for joining the Army and the soldier's perception of his/her job (MOS). These items were included because it was fel' that they related to the dimension of job satisfaction and thus, during the data analysis phase of the project, a job satisfaction measure could be developed from them. The second ARI-provided set of questions was called the Job History and Status report and contained 33 questions about the soldier's personal characteristics, job history, education, training and physical fitness. This second set of questions was included to provide ARI with data for an in-house study of potential Army-wide performance measures. Therefore, it has no direct relevance to the interest survey development effort.

Study Participants

Data were collected from a total of 527 enlisted personnel, representing 29 separate units within the 193rd Infantry Brigade. These individuals came from a total of 15 different MOS, however only five MOS were represented by a significant number of individuals: infantry (MOS 11B), mechanics (MOS 63B), drivers (MOS 64C), medics (MOS 91B), and military police (MOS 95B). Almost 94 percent of the sample were males.

Developing a Job Satisfaction Index

The plan for the analyses of the interest survey triad results specified that the analyses would be carried out within a subset of the troops from each MOS who were satisfied with their jobs in the Army. Since there are no regularly collected satisfaction indices for Army troops, eight items thought to relate to job satisfaction and job performance had been included in the data collection booklet at the suggestion of ARI staff.

In order to develop a job satisfaction index from these eight items, three iterated principal axis factor analyses (a one factor, a two factor, and a three factor solution) using squared multiple correlations as the initial communality estimates were carried out. The principal axis solutions were rotated to orthogonal simple structure matrices using the Varimax procedure and these Varimax matrices were further rotated to oblique simple structure solutions using the Promax (Procrustes) procedure. Of the three factor analyses, the three factor solution most clearly extracted a factor (Factor 3) that could be identified as job satisfaction. (Factor 1 appears to be a global factor related to characteristics of the job, and Factor 2 appears to be self-evaluation of job performance.)

The simple structure matrix vector for the third factor was then converted to a scoring coefficient vector, as would be used to actually produce a score on job satisfaction for an individual.

From this scoring coefficient vector it was clear that Item 7 was the only item making any significant contribution to the job satisfaction index as defined by factor analysis. Accordingly, it was decided to use Item 7 as the index of job satisfaction and not to include the responses to any of the other seven items. Item 7 was a simple question which asked the the soldier to use a four point scale to respond to the statement "I enjoy doing the type of work my job requires." Individuals who responded either 1 or 2 to Item 7 were classified as satisfied while individuals who responded either 3 or 4 were classified as dissatisfied.

Selecting Triads to Retain

The project staff had initially planned to develop scales to measure interests in seven of the major Army high density MOS clusters:

Armor
Clerk
Cook
Electrician/Electronics
Infantry
Mechanic
Medic/Lab Technician

Because no troops from four of the seven target MOS clusters were available, it was not possible to employ the strongly empirical approach to identifying discriminating triads that had been planned. Instead, a modified approach, which the project staff called a "rational-empirical" approach, was adopted to identify the triads that would be retained for inclusion in the final version of the interest survey.

This so-called "rational-empirical" approach was implemented in the following way. For each of the five MOS clusters for which troops from the 193rd Infantry Brigade were tested, and for the four additional target MOS clusters from which no troops were tested, the project staff made a rational judgment with regard to which options of which triads should discriminate which MOS clusters. In other words, based on their knowledge and experience, the two senior project staff members made an "educated guess" about which options of which triads would be chosen more or less often than the average by individuals in the nine MOS clusters. That is, which triads would discriminate between which MOS clusters. Then, for the five MOS clusters for which responses to the triads were available, an empirical determination of the discriminating power of the triads was made. For these five MOS clusters the results from the empirical determination were compared with the rational judgments of the project staff to provide an estimate of the validity of the judgmental process used by the project staff. While no formal, numeric validity coefficient was calculated, the level of agreement between the rational and empirical estimates of the discriminating power of the triads for which empirical data were available was sufficiently high to cause the project staff to feel that their judgments for the other four MOS clusters would be useful in the final triad selection process. Overall, the empirical results indicated that about eighty percent of the triads that the project staff identified were in fact discriminating in the anticipated direction.

The final 100 triads were selected in the following manner. Fifty-eight triads were selected because they met two empirical criteria: for the total set of 225 triads, they were in the top 100 triads on the basis of an overall discrimination index; and they were also in the top 100 triads for at least seven of the ten rankings of discrimination indices for pairs of MOS. There were no triads that were in the top 100 for at least seven of the ten rankings that were not also in the top 100 on the basis of their overall discrimination indices. The other 42 triads were included in the final set because, in the opinion of the project staff, they showed the greatest promise of discriminating individuals whose interest patterns would tend to place them in one of the target MOS clusters for which no empirical data were available. It would be possible to quibble with the selection of some of the triads that were included in the final set of 100, and in fact the project staff spent a great deal of time discussing the selections. However, the project staff feels that in the absence of adequate empirical data, this set of 100 triads is as defensible a set of 100 triads as can be selected and is more defensible than most sets. To create the ARIIS, the final 100 triads were ordered according to the length of the longest option in the triad.

REFERENCES

Clemans, William V. An index of item-criterion relationship. Educational and Psychological Measurement, 1958, 18, 167-172.

Kuder, Frederic Activity Interests and Occupational Choice. Chigago: Science Research Associates, 1979.

Kuder, Frederic, and Diamond, Esther, E. Kuder DD Occupational Interest Survey: General Manual (Second Edition). Chicago: Science Research Associates, 1979.

PREDICTIVE VALIDITY OF SURROGATE MEASURES OF JOB PERFORMANCE FOR NAVY GENDETs

by

Charles H. Cory, Navy Personnel Research & Development Center

Like many other organizations, the Navy for a number of years has been wrestling with problems of validating personnel selection tests against job performance criteria. In this connection, the Navy, and particularly the Navy Personnel Research and Development Center, the lead personnel research center for the Navy, has carried out a number of studies using supervisory marks as job performance criteria (Cory, 1976; Cory, Neffson, and Rimland, 1981; Cory, 1982). Refinements of the supervisory evaluation scale have been used (Borman, Toquam, & Rosse, 1979) and a survey of the job performance field to determine the current state-of-the-art has been carried out (Vineberg & Joyner, 1981). As a result of these studies we have concluded, perhaps not surprisingly, that supervisors' marks are not useful as performance criteria for the bulk of Navy jobs. Validity coefficients for supervisors' marks are simply too low and too unreliable for the marks to be useful as performance criteria for most jobs.

One choice resulting from this conclusion is between job knowledge and job sample tests as measures of performance, both of which alternatives are costly and time consuming to develop and to utilize. However, while developmental effort with job knowledge and job sample tests is being carried on, another choice is to investigate as a performance criterion measures which are available on a routine basis, are capable of being collected and processed inexpensively, and have high face validity. I am referring to surrogate measures of job performance--criteria derived from the operational records which measure outcomes related to and/or based on job performance. For instance, categories such as job level attained and speed of advancement are obviously related to job performance, although they are not direct measures of job performance such as supervisors' marks are purported to be (Vineberg & Joyner, 1981).

An exploratory study of the characteristics of surrogate job performance criteria was conducted at NPRDC using General Detail personnel (GENDETs). GENDETS are enlisted personnel who do not receive Navy technical school training for specific jobs. Instead they are sent directly to the Fleet and serve there in apprenticeship positions, generally in the Seaman, Fireman, and Airman ratings, where they receive training on the job. Because these positions are separate from the school training pathway of Navy advancement, measures of job performance are the only useful selection criterion to use for GENDETs.

For this study, a data set was formed by extracting from 10 data bases the records of first-term, non-prior-service male GENDETs who were enlisted as Els. The 10 data bases had been used for predictive validation studies which had been conducted at NPRDC from 1968 through 1976. Each of the studies had utilized as predictors operational variables formed not only from the classification test scores available operationally, but also biographical variables, and a set of experimental predictors.

The 10 data bases are shown in Table 1. Five of them were from research conducted in connection with Project 100,000. In addition, data bases used for the predictive validation of the Navy Vocational Interest Inventory, the Gates-McGinity reading test study and the BCS/Cleff Study were used, plus two

small studies, Computerized Perceptual Tests and the Technical Classification Assessment Center Study.

Predictors

Operational predictors included four biographical variables: Age, Years of Education, AFQT and Success Chances for Recruits Entering the Navy, (SCREEN) a composite variable based on years of education, age, marital status, and AFQT mental group. SCREEN previously had been found to be effective as a predictor of Navy enlisted personnel attrition. Scores for the Basic Test Battery and the Special Classification Tests, a total of eight tests which were the forerunners in the Navy of the ASVAB tests and measured about the same mental abilities were also used as predictors.

A total of 114 experimental predictors were formed for use in the seven data bases used for the predictive validity analyses, an average of 16 per data base. These included 17 aptitude measures, the majority of which were developed as culture-fair measures for the Project 100,000 studies, and 75 predictive scales developed from three biographical information questionnaires and two vocational interest tests. The other 22 experimental predictors consisted of 13 vocational interest scales, 3 achievement test scores and 6 miscellaneous measures.

Criteria

The source of criterion data was the Navy Enlisted Cohort History (NECH) tape, originated and maintained by the Naval Health Research Center. The NECH contains comprehensive data on the career histories of all enlisted personnel who have entered the Navy since 1 January 1965. From it four career history variables were extracted to serve as surrogate job performance criteria. They were (1) Rated/Non-rated, a binary variable coded "1" if rated and "0" otherwise, (2) Days to E4, the total days elapsing between enlistment into the Navy and achievement of E4 status, (3) Highest Pay Grade, the highest pay grade achieved by the individual during his Navy career, and (4) Disciplinary Record, a weighted composite formed from the total unauthorized absences, desertions and demotions. This variable is negatively scaled. Zero, indicating 0 disciplinary infractions is the highest score. A global supervisory performance evaluation mark which was present on five of the data bases was used to form the fifth criterion. It was: (5) Overall Performance, a supervisors' global evaluation of job performance recorded on a 5-point Likert scale, ranging from "1", Lowest 20%, to "5", Highest 20%.

RESULTS

The following types of analyses were carried out for the five criteria: (1) Criterion Intercorrelation Analyses, (2) Means for the six AFQT mental level groups (1s, 2s, Hi-3s, Lo-3s, Hi-4s, and Lo-4s) and (3) Predictive Validity analyses.

Criterion Intercorrelations

Table 2 shows the Pearson Product Moment intercorrelations of the five criteria and AFQT. Because of the very large sample sizes, all but one of the correlations are statistically significant even though many of them are too small to be of practical significance. In general, except for the high relationship between Rated/Non-rated and Highest Pay Grade ($\underline{r}$ = .71) the four career outcome

variables were relatively independent with three of the five other correlations being .10 or less. Overall Performance (OVER), the supervisory evaluation, was not highly related to the surrogate job performance measures (maximum $\underline{r}$ = .17). On the other hand, AFQT, a measure of general mental ability, was substantially related to three of the four career outcome criteria, ($\underline{rs}$ = -41, .35, and .25) although it had very low relationship to OVER ($\underline{r}$ = .06

Mean Performance of AFQT Mental Level Categories

Table 3 elaborates on the relationships shown in Table 1 by presenting means for the five criteria for personnel in six AFQT mental level groups. Comparison of the criterion means illustrates that for some criteria, the magnitude of the effects of AFQT on performance were very large. For instance, the percentage of Mental Group 1s who became rated was nearly four times that of Hi-4s, nearly twice that of Lo-3s, and 50 percent greater than Hi-3s. Similar large differences are evident in days to E4. In contrast the associations of mental level with changes in OVER and Disciplinary Record were not very pronounced.

The percentages of mental level groups achieving E3 and E5 pay grades and the days required to achieve these pay grades were also computed and were used to compute the percentages of days in a 4-year enlistment period which would have been spent at or above a particular pay grade. These statistics are shown in Table 4.

There was a low positive relationship between mental level group and the average percentage of time spent at E3 level or higher. In contrast, a substantial monotonic increasing relationship existed between mental level and percentage of time spent as E4s and E5s. Personnel with higher mental abilities spent a much higher percentage of their overall enlistment period in rated status than did personnel with lower mental abilities. For instance, Mental Group 1s, on average, spent 46% of their enlistment period as rated personnel, compared with 12 percent for Lo-3s, 2% for Hi-4s and 0% for Lo-4s. If it is assumed that time spent as a rated person is of greater value to the Navy than time spent on the job in a partial performance or training capacity, these data clearly demonstrate the importance of mental level for the selection of enlisted personnel for GENDET assignments.

The predictiveness of each of the criteria using predictive composites formed from up to five variables was also investigated. For this step the seven large data bases in the study were split in half and predictive composites were formed using a step-wise multiple regression program. The variables in the predictor composites thus selected were given unit weights and the predictor composites formed were used to compute cross-validation coefficients on the holdout samples and back validation coefficients on the predictor selection samples. Average cross-validation coefficients and standard deviations for each criterion were computed using a validity generalization formula recommended by Schmidt and Hunter (1977), and difference scores for each criterion were computed by subtracting the average cross-validation coefficient from the average back validation coefficient. These statistics are shown in Table 5.

It is clear from Table 5 that both Highest Pay Grade and Rated/Non-rated were much more predictable than were supervisors' marks. Adjustment of the RAT/NR coefficient to compensate for the restriction in magnitude occurring because RAT/NR was a binary variable (not shown) indicates that the coefficients for HIPG and RAT/NR were equal. These coefficients indicate that there was nine times more predictable variance for HIPG and RAT/NR than for OVER.

The stability of the predictive validity coefficients is shown by the small average shrinkages occurring from the Back Validation to the Cross-Validation mean coefficients. Shrinkage of OVER on cross-validation (.05) was considerably greater than that of any of the career outcome criteria.

Follow-up Research

We consider the findings to be promising, and NPRDC has undertaken additional studies to evaluate the use of career outcome criteria for all enlisted personnel. One of the studies will investigate the feasibility of developing a job performance weight for inclusion in the Navy CLASP system so that job performance may be considered in addition to achievement in Navy technical schools when assigning enlisted personnel.

REFERENCES

Borman, W. C., Toquam, J. L., & Rosse, R. L. An inventory battery to predict Navy and Marine Corps recruiter performance: Development and Validation (NPRDC Technical Report 79-17) (AD-A069-371). Unclassified.

Cory, C. H. A comparison of the job performance and attitudes of Category IVs and I-IIIs in 16 Navy ratings (NPRDC Technical Report 76-35). San Diego: Navy Personnel Research and Development Center, May 1976. (AD-A024-642)

Cory, C. H. Fleet follow-up of personnel appraised in a technical classification assessment center: Pilot study (NPRDC Technical Note 82-23). San Diego: Navy Personnel Research and Development Center. July 1982.

Cory, C. H., Neffson, N. E., & Rimland, B. The validity of a battery of experimental tests for predicting performance of Navy Project 100,000 personnel (NPRDC Technical Report 80-35). San Diego: Navy Personnel Research and Development Center, September 1980. (AD-A035-744)

Schmidt, F. L. & Hunter, J. E. Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 1977, 62, 529-540.

Table 1

Data Bases Used

Name of Data Base	Year Collected	Description	N	N (AFQT
100K Research, Phase 1	1967	Experimental Tests for Project 100K	2,898	2,566
100K Research, Phase 2	1968	Experimental Tests for Project 100K	6,245	4,850
100K Research, Phase 3	1968	Experimental Tests for Project 100K	5,992	4,589
100K Research, Phase 4	1969	Experimental Tests for Project 100K	6,120	4,521
100K Research, Phase 5	1970	Experimental Tests for Project 100K	3,687	3,287
Navy Vocational Interest Inventory	1969,1970	NVII as predictors of A-School grades	17,109	15,041
Computerized Perceptual Tests	1972	Computerized tests as predictors of supervisors marks	83	83
Gates-McGinity Reading Test	1974,1975	Study of reading level of Navy enlisted input	2,761	2,754
Assessment Center Study	1975	Hands-on performance tests for classifying GENDETs	86	86
BCS/Cleff Study	1976	Interest Inventory for A-School selection	1,250	1,243
TOTAL			46,231	39,020

Table 2
Intercorrelations of Five Criteria & AFQT

	DAE4	HIPG	DISR	OVER	AFQT	MEAN	SD	N
Rated/Non-Rated	--	.71***	-.08***	.09***	.25***	.44	.50	46,231
Days to E4*		-.17***	.10***	-.02	-.41***	873	440	20,159
Highest Paygrade			-.14***	.17***	.35***	3.62	1.21	46,231
Disciplinary Record*				-.07***	-.06***	.75	1.60	46,231
Overall Performance					.06***	3.13	.26	5,625
AFQT						56.43	20.85	39,020

* Variable was negatively-scaled
***p<.001

Table 3
Mean Job Performance Indices of GENDETs
Classified by Mental Level

	Mental Level Group						
Criterion	1	2	Hi-3	Lo-3	Hi-4	Lo-4	N
% Achieving Rated Status	77	62	51	42	20	73	39,019
Days to E4	580	726	929	1,026	1,301	1,487	19,487
Highest Pay-grade	4.75	4.23	3.78	3.42	3.04	4.65	39,019
Disciplinary Record	.55	.66	.85	.99	.53	.83	39,019
Overall Performance	3.16	2.88	2.73	2.58	2.75	2.43	3,712

Table 4

Average Percentage of Days Worked at or above each Pay Grade Level during a Four-Year First Enlistment Period

	Mental Level Category						
	1 N=1629	2 N=11893	H3 N=10332	L3 N=11350	H4 N=3522	L4 N=293	Total Group N=39019
E3	72	67	60	51	48	67	59
E4	46	31	18	12	2	-	20
E5	9	3	-	-	-	-	1

Table 5

Descriptive Statistics for Predictive Validities

	Average Cross-Validation	SD of Average CV Coefficient	Average Shrinkage on Cross-Validation
Rated/Non-rated	.35	.06	.01
Days to E4	-.29	.10	.03
Highest Pay Grade	.43	.08	.01
Disciplinary Record	-.13	.04	0
Overall Performance	.15	.09	.05

ABSTRACT

MILES Air Ground Engagement Simulation/Air Defense Training Effectiveness Analysis

Dale M. Dannhaus, PhD
Charles R. Hughes
Major John M. Shea

US Army TRADOC Systems Analysis Activity

MILES employs an eye-safe laser beam to simulate a firing weapon and laser detectors attached to targets to assess casualties. The MILES program allows two-sided force-on-force free-play training exercises. MILES AGES/AD is presently being developed to expand the MILES system to include helicopters and division short range air defense weapon systems in combined arms tactical training exercises. This presentation highlighted the results of the training effectiveness data collected during Aug-Oct 1981. Three groups participated in the test, each consisting of one attack helicopter platoon (5 AH1S; 3 OH-58; 1 UH-1), one CHAPARRAL section, one VULCAN section, three STINGER teams, one M113, and three M60A1 tanks. A video cassette of the exercises was shown. The presentation provided an assessment of the potential collective training value of the MILES AGES/AD devices.

AD P000859

THE CODAP80 "RANDOM" PROCEDURE

Richard W. Dickinson

Occupational Research Division
Industrial Engineering Department
Texas A&M University

INTRODUCTION

CODAP80, a new job analysis software system, is being developed that will allow analysts greater flexibility in the way occupational information can be manipulated and displayed. This paper discusses the incorporation of a procedure into CODAP80 that will allow the convenient determination of the effects of different sampling distributions on the results derived from occupational information.

CODAP80

At present, a new version of job analytic computer software is being developed, called CODAP80, that will greatly increase the job analysts' ability to answer questions of occupational information. In this new system, users access the occupational database through the use of an easy to learn, English-like language. There are no restrictions on data access and retrieval, allowing any piece of information residing on the database to be processed. The basic design philosophy of the new system is to conceptualize job analytic database information as a two-dimensional matrix in which incumbents represent the columns of the database, and the variables the incumbents are measured on representing the rows of the database (see Table 1).

Using CODAP80, job analysts will have the ability to perform calculations on any variables (rows) in the database across any subset of incumbents (columns). The resultant calculations may then be added to the database for further processing. The flexibility of the new system allows the added convenience of "symmetry," in which any calculations performed across database columns may also be performed across database rows. For a more in-depth discussion of the new system's operational capabilities and characteristics, the reader is referr · to Dickinson (1979, 1980).

The Random Procedure

The purpose of the RANDOM procedure is to give the job analyst a convenient method by which database columns or rows may be randomly selected for processing. Since a CODAP80 database may contain virtually any information, the applications of the procedure are as various as there are types of occupational data. For example, were a CODAP80 database to contain items from, say, a test bank of questions, the RANDOM procedure could be used to

provide a random listing of a subset of the total for purposes of educational assessment or selection. To aid in readministration, the RANDOM procedure will optionally save the aggregate of rows (called, in CODAP80 terminology, a "module") that was randomly selected for reference at a later time.

The RANDOM procedure is particularly convenient and easy to use. For example, suppose a user desired to randomly select 100 incumbents from a study population for processing. The user would only need to code:

```
RANDOM COLUMNS INCUMBENTS 100
  RANDINCS '100 RANDOMLY SELECTED INCUMBENTS'.
```

In the above CODAP80 syntax, the keyword RANDOM identifies the command. COLUMNS indicates to the procedure that the selection is to be made from the columns of the database (specifically, from the incumbent columns of the database, as indicated by the INCUMBENTS keyword). The integer number following the INCUMBENTS keyword indicates how many of the columns to select. RANDINCS is the user supplied ID that will be permanently associated with the randomly selected "group" of incumbents and the character string enclosed in single quotes that follows the ID is the user supplied descriptive information that will be stored along with the ID. From this point, the user need only to refer to the ID RANDINCS in other procedures for the CODAP80 system to know what incumbents were randomly selected.

Hierarchical clustering can be, depending on the number of objects being clustered, very expensive in terms of computer processing. For purposes of classification it may well be that clustering a randomly selected subset of the incumbent population of interest could yield patterns of common time spent that would serve as well as a cluster solution derived from the total population. The RANDOM procedure used in conjunction with CODAP80's CLUSTER procedure would provide an easy method of determining this. The flexibility of the RANDOM procedure even allows a subset of incumbents to be randomly chosen from <u>within</u> a cluster group.

As an alternative to "pure" random sampling, the RANDOM procedure provides the user with the "KTH" option. This option directs the procedure (through user input) to select every "kth" element of a population, with the first element being chosen randomly.

CONCLUSION

The CODAP80 RANDOM procedure provides the user with an easy and convenient way to investigate sampling effects on the overall results derived from occupational information. Along with CODAP80's other procedures for data manipulation and display, the RANDOM procedure allows job analysis to move away from single study investigations to more general database applications.

TABLE I

CONCEPTUAL REPRESENTATION OF JOB ANALYTIC DATABASE

	I1	I2	I3	I4	I5	I6	I7	I8	I9
H1									
H2									
T1									
T2									
T3									
T4									
E1									
E2									
E3									
E4									

H = History or Background Information
T = Task Information
E = Equipment Usage
I = Incumbent

REFERENCES

Dickinson, R. W. The New Codap System -- Design Concepts and Capabilities. Paper presented at the Military Testing Association annual meeting, San Diego, California, 1979.

Dickinson, R. W. The New Codap System's Enhanced Hierarchical Clustering Capability. Paper presented at the Military Testing Association annual meeting, Toronto, Canada, 1980.

PREDICTING ATTRITION IN THE ARMY INITIAL ENTRY ROTARY WING COURSE

John A. Dohme, William R. Brown and Michael G. Sanders
US Army Research Institute Field Unit, Fort Rucker, AL

Selection testing for Army flight training goes back to the days of the Army Air Force in World War II and the august crew of psychologists who were called upon to serve their country in time of war. A group including J. C. Flanagan, Neal Miller, Paul Fitts, Edwin Fleishman, Arthur Melton and others launched their successful careers developing tests and measures to select and classify aviators. The Flight Aptitude Selection Test (FAST), which is currently in operational use, has its development (and a few of its items) directly traceable to that research effort undertaken during World War II. Parenthetically, we can note that a few of the aforementioned psychologists are also currently in operational use.

Selection procedures, especially those rooted in antiquity, benefit from an occasional reevaluation and/or revalidation since there are periodic changes in the flight training curriculum and also drifts in the qualifications of the applicant pool. Since the initial development of the Army Air Forces Qualifying Examination (AAFQE) in 1942, the Army has developed the helicopter as a tactical vehicle and weapons platform. In addition, the Army has initiated the aviation warrant officer training program recently ~~heralded~~ in TV spots offering "High school to flight school" training. ~~Thus,~~ Our research program at ARI continuously evaluates the selection, mission assignment and training of both commissioned and warrant officer aviators. The Aviation Center's 36 week Initial Entry Rotary Wing (IERW) training course graduates combat ready aviators who have been tactically trained in either the Aeroscout or Utility mission. This paper reviews recent research aimed at optimizing selection in order to minimize attrition in IERW training.

5th

Historically, selection of Army aviators began with the efforts of COL Flanagan's group during WWII. Their AAFQE reduced the attrition rate from 75% with unselected trainees to 35% (Davis, 1947). In the current IERW program, attrition is approximately 7.2% for commissioned officers with about 50% of that attrition occurring because of flight deficiencies. Among the Warrant Officer Candidates (WOCs), overall attrition is approximately 20.5% with 14% of that attrition related to flight deficiencies. Part of the discrepancy in flight deficiency attrition rates relates to the fact that commissioned officers have been through officer development training before coming to the IERW training program, whereas the first 6 weeks of the WOC training program is Warrant Officer Candidate Military Development (WOCMD) training. Thus, over half of the WOC eliminees have attrited before flight training begins. However, the WOC attrition rate, looking only at individuals who have successfully completed WOCMD, is still 15.6%, double the rate for commissioned officers. Although these attrition rates are rather low vis a vis other flight training programs, with IERW training costs running approximately $125,000 per student, there is continuing interest at Fort Rucker in minimizing attrition and optimizing selection.

METHOD

In FY 82, the present authors reviewed the causes and correlates of attrition in the IERW course for all trainees in FY 80 and the first half of FY 81

(Dohme, Brown and Sanders, 1982). In all, the training records of 3,293 flight students were reviewed; 1,108 commissioned officers and 2,185 WOCs. Each student's progress through the course was tracked (including medical or administrative leave time and "turnbacks" to an earlier training class) until either graduation or elimination. Eliminations were analyzed in terms of the stated reason for the elimination, the training phase during which elimination occurred, the incidence of single or multiple turnbacks, and the race of the eliminee. These analyses did not shed much light on the attrition process except in showing no clear differences between black and white eliminees.

Training records were searched for variables that might be predictive of IERW training performance. Since the FAST and the General Technical (GT) subtest from the Armed Services Vocational Aptitude Battery (ASVAB) are prerequisites for application to flight training, they were obvious candidates for predicting graduation/elimination. Other potential predictors included in the analysis were the Skills Technical (ST) subtest from the ASVAB, age at the time of IERW course matriculation, and amount of formal education (where 12 years equates to a high school diploma).

The predictor variables were related to the criterion variable individually (using biserial correlation) and in combination (using discriminant analysis). Two methodological limitations should be noted in this approach. First, the GT and FAST scores have been used administratively to screen the individuals who enter flight training. This reduces the range of observed scores on the GT to approximately the top 35% of the population and on the FAST, to approximately the top 50% of the WOC population and the top 92% of the officer population.* In addition, the criterion measure reflects components other than the individual's ability to master the flight training tasks. Overall, 26.5% of officer and WOC eliminations are related to flight deficiencies while the remainder are related to medical problems, administrative problems (such as illness in the family), resignation and lack of military development.

RESULTS AND DISCUSSION

Figure 1 presents biserial correlations of the predictor variables with the criterion (graduation/elimination).

	OFFICERS		WOCs	
PREDICTOR VARIABLE	BISERIAL r	SIGNIFICANCE	BISERIAL r	SIGNIFICANCE
GT	Not Applicable		.07	NS
ST	Not Applicable		.13	$p<.05$
EDUCATION	.18	$p<.01$	-.08	NS
AGE	-.46	$p<.01$	-.36	$p<.01$
FAST	.32	$p<.01$	.26	$p<.01$

Figure 1. Biserial Correlations of Predictor Variables with Graduation/Elimination for Officers and WOCs.

*Except for a 9-month period during FY 80 when the WOC FAST cut score was lowered from 300 to 270 corresponding to approximately the 34th percentile.

The biserial correlations are presented separately for officers and WOCs because the two groups formerly took different forms of the FAST. Presently, both applicant groups respond to the same form of the Revised FAST (RFAST) which was first fielded in mid FY 80.

Another way to consider the prediction of IERW training performance is to plot the percent of students graduating (also interpretable as the probability of graduation) as a function of scores on each predictor variable. Figures 2-7 present these data for the variables GT, ST, years of education, age at entry and FAST score for WOCs and officers.

To evaluate race as a predictor, we performed a stepwise discriminant analysis on the data to classify students as probable graduates or eliminees. After the other predictor variables were entered into the stepwise discriminant procedure, race was forced in last. The rationale was that if race adds predictive efficacy after inclusion of the traditional predictor variables, then there are performance differences associated with race that are not accounted for by the other predictors. This outcome would signal a problem with unfairness in the predictor variables and/or racial bias in the IERW training program. The F to enter values for the stepwise discriminant function coefficients are presented in Figure 8.

OFFICERS			WOCs		
VARIABLE	F TO ENTER	SIGNIFICANCE	VARIABLE	F TO ENTER	SIGNIFICANCE
AGE	22.83	$\rho<.01$	AGE	23.53	$\rho<.01$
FAST	8.31	$\rho<.01$	FAST	13.53	$\rho<.01$
EDUCATION	.12	NS	EDUCATION	9.83	$\rho<.01$
RACE	.06	NS	RACE	.25	NS
			GT	.01	NS

Figure 8. Significance of Variables Entering into the Stepwise Discriminant Analysis.

As the F values demonstrate, race is not a significant predictor. In fact, race adds virtually no information to the prediction of graduation/elimination once the other predictive relationships have been accounted for. Moreover, the univariate F ratios for race in the discriminant analysis do not reach significance. For officers, the univariate ratio is $F = .46$ ($\rho = .50$) and for WOCs, $F = 2.01$ ($\rho = .16$). Thus, race is not significantly related to IERW training performance and we may conclude there is no observed racial effect in the prediction of graduation/elimination in the IERW training program.

Figure 2 shows that the GT subtest is not an effective predictor of IERW graduation/elimination in the range plotted. Since the GT is used to screen individuals for acceptance into the flight training program, there are no scores below 110 in the trainee population. This truncation in range probably affects the predictive relationship to lower the apparent effectiveness of the GT. Figure 3 demonstrates that the ST subtest is somewhat more effective as a predictor than the GT. However, the range of scores is not as greatly restricted on the ST subtest (see Figure 3) since it is not currently used for selection. Also, the intercorrelation between the two subtests is $r = .69$. Subsequent research will evaluate the ST subtest as a selection

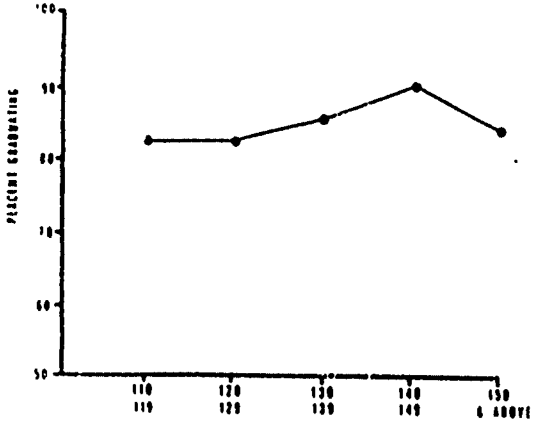


Figure 2. GT for WOCs as a Predictor of IERW Graduation/Elimination.

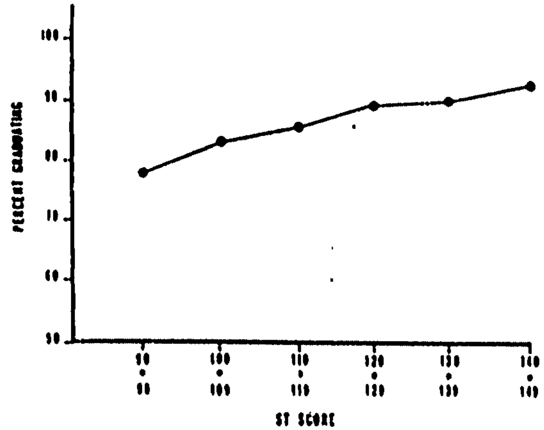


Figure 3. ST for WOCs as a Predictor of IERW Graduation/Elimination.

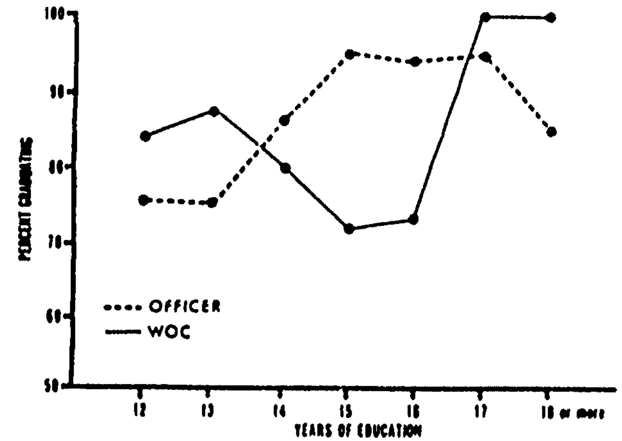


Figure 4. Education for Officers and WOCs as a Predictor of IERW Graduation/Elimination.

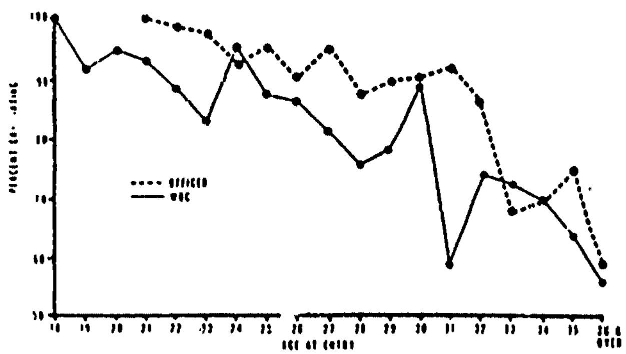


Figure 5. Age for Officers and WOCs as a Predictor of IERW Graduation/Elimination.

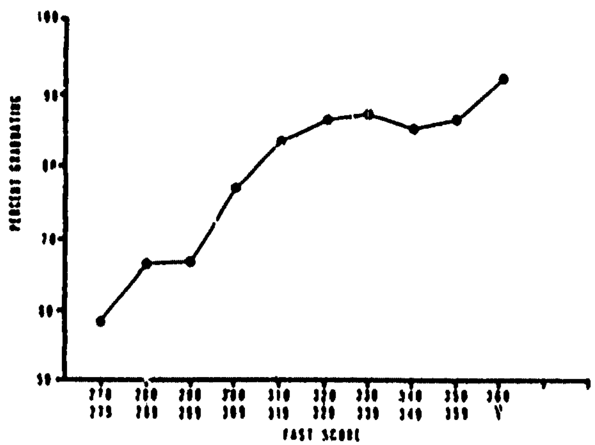


Figure 6. FAST for WOCs as a Predictor of IERW Graduation/Elimination.

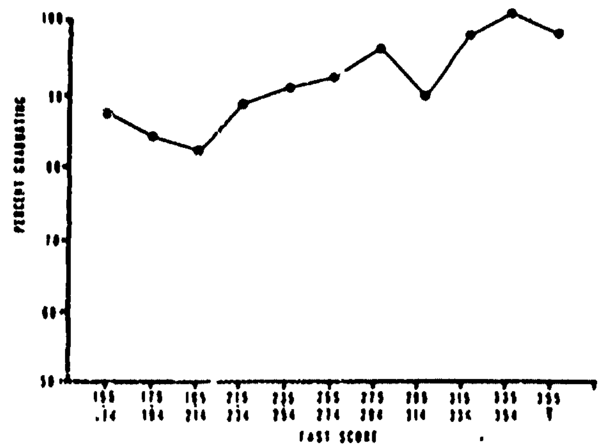


Figure 7. FAST for Officers as a Predictor of IERW Graduation/Elimination.

test to determine whether it adds to the prediction of WOC performance over and above the variance accounted for by the GT subtest.

Figure 4 shows that education has a complex relationship to the probability of graduation for WOCs and officers. Figure 1 reflects that the biserial correlation of education and the criterion is significant for officers (r = .18) but not for WOCs (r = -.08). Figure 7 shows that education adds significantly to the stepwise discriminant analysis but only for WOCs. Drawing conclusions regarding the relationship between education and graduation must be tempered by the fact that some points on the graph in Figure 4 represent very small sample sizes. For example, 7 WOCs have 17 years of education and 4 have 18. Therefore, more research should be performed to evaluate education as a predictor of graduation/elimination.

Figure 5 shows that age is closely related to the probability of graduation/elimination for both WOCs and officers. While both curves show perturbations due to the relatively small sample sizes at each age level, there is a linear trend from age 18 to 30 with an inflection downward past age 30. The biserial correlation and discriminant analysis results show that age bears a strong inverse relationship to graduation from the IERW course.

Figures 6 and 7 show the FAST as a predictor of graduation/elimination for WOC and officer students. A comparison of the 2 figures shows that the FAST is a more effective predictor for WOCs than for officers. Additionally, the WOC FAST battery is most effective as a screening test in that Figure 6 shows a steeper slope for lower scorers than for higher scorers. In other words, the FAST can identify those individuals who are greater risks in IERW flight training.

The best use of this predictor information is to combine the predictors in a stepwise fashion to optimize selection. In fact, we're currently developing a selection procedure for Warrant Officer Branch of MILPERCEN using discriminant analysis to combine the predictor scores discussed above with judgmental scores from the selection board members. The judgmental scores include fresh information in the selection algorithm such as the applicant's aviation background, military experience, and letters of recommendation. Optimal selection of applicants can be achieved by developing and cross-validating an algorithm that uses the variables discussed above as well as RFAST <u>subtest</u> scores, each with its appropriate β weight.

Recent research with the Revised FAST (RFAST) by Lockwood (1982) demonstrated greater predictive validity using the 7 subtest scores in place of the composite score. Eastman and McMullen (1978) estimated that the predictive validity for the FAST was r = .38 for WOCs and r = .44 for officers. The use of RFAST subtest scores in Lockwood's multiple regression equation raised the validity estimates to R = .42 for WOCs and R = .56 for officers for a sample of 108 student pilots. While this finding is subject to cross-validation, it suggests the utility of combining subtest scores in the optimal WOC selection algorithm. In fact, when Lockwood included the ST score along with RFAST subtest scores, the predictive validity for WOCs was raised to R = .68.

Research is currently addressing a number of related selection and assignment issues. An alternate form of the RFAST is being tested for

equivalency at the current time. The RFAST has been evaluated for bias and replacement items have been developed to substitute for those found to be biased for and against minority groups. A front-end analysis has been completed identifying the abilities required to fly each of the Army helicopter missions, Aeroscout, Attack, Cargo and Utility. Work is underway to develop tests and measures that will permit differential assignment of student pilots to specific mission training as part of the IERW curriculum. At the same time, an ability analysis has been performed on the phases of IERW training where most flight deficiency attrition occurs, primary and instruments. New FAST subtests will be developed to measure these critical abilities and, hopefully, reduce attrition. Since over half the WOC attrition occurs in WOCMD, a study is being conducted to develop predictors that will identify applicants who are likely to be eliminated in that training phase. In short, we're working the problem and we think COL Flanagan would enjoy being a part of our research effort.

REFERENCES

Davis, F. B. (Editor) The AAF qualifying examination, Army Air Forces - Report #6, Washington, DC, 1947 (AD 651 782).

Dohme, J. A., Brown, W. R., and Sanders, M. G. Comparison of minority/majority attrition in the Army initial entry rotary wing aviator course: Predicting graduation/elimination from FY 80 and 81 entry scores. Research report, Army Research Institute Field Unit, Fort Rucker, AL, 1982.

Eastman, R. F. and McMullen, R. L. The current predictive validity of the flight aptitude selection test. Research memorandum 78-2, Army Research Institute Field Unit, Fort Rucker, AL, 1978.

Lockwood, R. E. Evaluation of the flight aptitude selection test. Technical report, Army Research Institute Field Unit, Fort Rucker, AL, 1982.

AD P000861

THE GRADUATES' VIEW OF NAVAL ROTC

Andrew N. Dow, Ed.D.
Project Director, Naval Science Evaluation and Analys_
Chief of Naval Education and Training, Pensacola, FL 32508

I. INTRODUCTION

In response to a Naval Inspector General (IG) requirement the Chief of Naval Education and Training (CNET) developed, distributed, and analyzed the NROTC Graduate Feedback Survey. This survey looked at the noncurricular as well as curricular aspects of NROTC. The IG suggested that such a survey be modeled after those of the Naval Academy (USNA) and the naval Officer Candidate School (OCS). In 1979 a form was developed and used on the 1977 year group. USNA surveys their graduates 3, 7, and 12 years after commissioning; CNET decided to use these same time frames. As of 1981, the year groups 67, 68, 69, 72, 73, 74, 76, 77, and 78 have been surveyed and analyzed. In the groups surveyed in 1981, 36 percent of the forms sent to Navy officers and 57 percent of those to Marines were received completed.

The survey form was similar to the one used at USNA. There are demographic items, 25 items asking for the value of the program and its phases, and eight to rate the value of the program to aspects of personal development. The 33 items use a five point rating scale; three others are open-ended questions.

The responses are put on magnetic tape, then tallied and analyzed by MIISA. Currently two programs from the Statistical Package for the Social Sciences (SPSS), "Cross tabs" and "Frequencies" are used; each makes distributions. In addition, "Cross tabs" divides the data along two dimensions and calculates the number and percent in each cell, while "Frequencies" computes a mean scale value for each and draws histograms of the responses.

Printouts are checked for trends and atypical responses. Using the percent who marked the two upper cells, the items rating the preparatory value of the various aspects of the program were ranked within groups and for some combination groups. The items that rated the value of the NROTC Program to aspects of personal development were also ranked. Mean scale values of both sets of items are observed, then analyzed for long-term trends and irregularities in addition to their relative ranking.

II. PRELIMINARY ANALYSIS OF SIX YEAR GROUPS

USNR CNET Hq Det 110 analyzed 6 year groups (67, 68, 72, 73, 76, and 77). Using the percent that marked the two favorable categories they ranked survey items 11 through 31 (except 24, 25, 27, and 28) for each group and for all six together. The rank orders of the items for these groups combined are:

1. First class cruise (taken after junior year)
2. Third class cruise (after freshman year)
3. Second class cruise (after sophomore year)
4. Example set by staff
5. Leadership and Management training
6. Physical Fitness Program
7. Staff counseling
8. Orientation course

9. Military Drill and Ceremonies
10. Naval Science Laboratory
11. Military law instruction
12. Seapower course
13. Contact with unit CO
14. Naval Engineering course
15. Administrative Procedures training
16. Social events and activities
17. Naval Weapons course

They also ranked items 32 through 39 on a comparable basis; these deal with the value of the NROTC program to personal competencies. These ranks are:

1. Sense of responsibility
2. Leadership abilities
3. Self-confidence
4. Communications skills
5. Decisionmaking skills
6. Think and act under pressure
7. Ability to manage time
8. Analytic skills

From these rankings, we find that the cruises and the examples set by the staff were the aspects felt to be most valuable. The ranks earned by items 32 through 39 show that the six groups felt that NROTC had contributed the most toward building their sense of responsibility, leadership abilities, and self-confidence.

III. ANALYSIS BASED UPON NINE YEAR GROUPS

In 1981 surveys were sent to the 69, 74, and 78 year groups. The data from these were added to the previous, then tabulated and analyzed by MIISA.

A. Characteristics of the Responding Groups

Table 1 (immediately below and on the next page) presents data descriptive of the respondents of nine year groups, singly and combined.

YEAR GROUP	67	68	69	72	73	74	76	77	78	ALL
NUMBER	132	165	179	171	238	288	705	563	563	3005
NO. FEMALE	0	0	0	1	35	0	38	34	12	120
% SCHOLARSHIP RESPONDING	83	75	71	85	89	94	92	92	92	89
% SCHOLARSHIP GRADS IN YEAR GROUP	76	46	57	72	84	87	87	88	87	72
USMC %	23	18	8	17	20	16	22	19	18	19
AVIATORS % (USN & USMC)	40	35	36	40	44	49	29	29	31	34

NUC % (SUB & SURF)	2.3	3.0	0.6	1.8	0.4	4.2	2.4	0.7	9.9	5.9
STAFF CORPS %	17	27	22	17	27	12	15	18	15	18
TECH & ENG MAJOR %	41	38	35	36	32	36	35	43	48	39
WOULD CHANGE MAJOR %	67	70	69	63	71	65	75	79	74	72
% CAREER UNDECIDED	2	5	7	46	37	39	62	70	59	49
% 20 YEAR EXPECTATION	96	94	89	35	48	46	15	15	19	35
# 20 YEAR EXPECTATION	127	155	159	60	115	133	106	84	107	117/YR GRP

TABLE 1

In every year group, the percent of scholarship holders responding was larger than the percent of that class that had held scholarships. Also, the proportion of technical majors (including engineering) is consistently less than one half. Further, 2/3 to 3/4 would select a different major if they could. The percent expecting to have a 20-year career increases radically--16 to 43 to 93 as we go from 3 to 7 to 12 years after accession--but the numbers increase only slightly. The makeup of the responding group changes with the passage of time. This raises a question-- Are the changes in response pattern with increasing seniority real changes based upon experience, or are they only the result of the different makeup of the group responding?

B. Evaluation of the Features of NROTC

When the ratings are processed through the SPSS program "Frequencies" a mean response value is calculated for each item. Using the mean values combined, the same items that USNR Hq Det 110 used were ranked. The similarity of the two sets of ranks is striking. Here are the 17 items in order based upon the means of nine year groups; the ranks at the right are from the six groups studied by Hq Det 110.

Rank (9 YG)		Rank (6 YG-USNR study)
1.	First class cruise	1
2.	Third class cruise	2
3.	Second class cruise	3
4.	Example set by staff	4
5.	Leadership and Management	5

6.	Physical Fitness	6
7.	Staff Counseling	7
8.	Orientation course	8
9.	Military Drill and Ceremonies	9
10.	Naval Science Laboratory	10
11.	Sea Power - Maritime Affairs	12
12.	Military Law	11
13.	Administrative Procedures	15
14.	Social Events and Activities	16
15.	Naval Engineering Course	14
16.	Naval Weapons Course	17
17.	Contact with unit CO	13

Items 32 through 39 ranked exactly the same as they did for the 6 year groups. The sameness of the rankings using the two different, but overlapping, methods shows that either is a satisfactory system of analysis.

When comparing the ratings of each item over the years, it was noted that some were rated about the same with minor fluctuations. A few were found to be more valuable to recently accessed officers than to their seniors.

The two items that had the largest increase in perceived value were:

Contact with NROTC Unit CO
NROTC Staff Counseling

Others perceived to be more valuable to recent graduates than to the earlier ones (in the order in which they appear on the survey):

Value of NROTC in competing with non-NROTC officers
Naval Science Laboratory Sessions
Physical Fitness
Example Set by NROTC Staff Members
Social Events and Activities
Navigation course
Leadership and Management courses
Amphibious Warfare (Marines only)

Whether this indicates real improvement, changing attitudes, or the different composition of the groups cannot be determined from these data.

C. Analysis and Synthesis of Subjective Responses with Commentaries

On several occasions, officers on temporary assignment to CNET have read the forms, then categorized, summarized, and tallied the responses to the open-end questions. A good example of this is the 1969 year group from which we find that:

Thirty-eight percent stated that training was weak in administration, including leadership, management, and military law. Thirty-six percent asked for more training directed toward the shipboard officer duties and the "Real World," with some asking for "hands-on" training. While we can't know the

exact nature of these expressed desires, they are for specific training, the kind that is better left to SWOS and other postaccession schools.

Thirt seven percent stated that some instructors were weakly motivated. This deficiency may be the result of casual selection processes, of inadequate instructor training, or of insensitive leadership. If the Navy is interested in probing the possible causes of these perceived difficulties, further research is needed.

Seventeen percent commented that the NROTC courses were weak or outdated. The continuing process of course review takes care of this.

Fourteen percent were dissatisfied with summer training, but the favorable ratings that it got in the objective evaluation overbalances these comments. The unfavorable comments are likely based upon real problems, but as it has been several years since these problems occurred, specific difficulties and their cures need not be identified.

Twelve and eleven percent respectively commented that there was too little training in writing and too much required science. General training in writing is a university function, and specific naval writing should be taught in postaccession schools or special courses, not in NROTC. As for the excess of science--that is difficult to evaluate. It may have resulted from poor quality instruction and/or from a failure to show the student how the knowledge of science fits into the everyday life of a naval officer.

IV. FINDINGS

The responses are favorable; programs and relationships were seen as valuable. Many of the difficulties pointed out by early groups have been spotted through other channels and ameliorated.

All groups praised NROTC for bringing the midshipman into the Navy gradually and merging the military training into the university experience.

Some praised the quality of the NROTC instructors, while others felt that weak instructors were the main shortcoming of the program. Some spoke of the value of a commanding officer with strong leadership, others mentioned the dulling aspects of a commanding officer on his "grave yard" tour. The single feature of the NROTC program that is rated very valuable the most often is the first class cruise.

These polarities in the feelings toward several features or personnel point out that these are sensitive items and deserve attention. In other words, when these items are good they are a positive influence and should be maintained as they are; when they are weak, they should be brought into line quickly.

The respondents did not perceive the formal naval science courses as especially valuable. This is confirmed by the items about the value of the overall NROTC program to areas of personal development. Sense of responsibility and leadership abilities being the aspects to which NROTC had contributed the most.

Seventeen percent of the class of 69 felt that the courses were weak or outdated. Of the courses, Naval Weapons (Ship Systems II) is rated the lowest. It is being revised at this time; as it is a sophomore course, the effects of the revision will not be known for several years. In any event, the feelings about courses should be referred to the course coordinators.

The two specialized courses taught by and for the Marines are valued more highly than most of the Navy courses. The restricted student body, the specificity of these courses, and the career orientation of the Marine instructors may account for this acceptance.

V. RECOMMENDATIONS

To provide continuity, permanent data bank facilities, and the advantages of other personnel, the Training Analysis and Evaluation Group (TAEG) should be brought in as resource persons and consultants. While the present data processing arrangements are satisfactory, and the in-house handling has not yet imposed an undue burden upon CNET (Code N-122), the long-range program should benefit from the proposed interaction with TAEG. Before there is any formal tasking of TAEG, N-122 personnel, CNET Code 002, and members of the TAEG staff should discuss this survey and its future.

The validity of using the Graduate Feedback Survey only with those officers that are 3 years removed from their accession should be investigated. Present observations of the returns from those who have been commissioned 7 and 12 years seems to show that the composition of the responding groups changes markedly as the time after graduation increases. At present, no year group has been resurveyed. We do not know the reaction of a group to a resurvey, nor how comparable the two surveys will be. Neither can we say whether the change in response pattern across the year groups is real, an artifact caused by changes in the group composition, or simple fading of the memory.

While the present form may be adequate, there are changes that might improve it. The current form requires that the responses be coded onto special sheets for the keypunchers. If changes are made, they should include reformatting to allow direct keypunching to save time and reduce errors. It also should be reviewed with an eye toward evaluating content, specific wording, and overall length. Any change that would make it more attractive and look easier to answer and return should increase the percentage of returns, and improve the representativeness of the sample.

If the Navy perceives a strong need for information that is not on the current form, then it should be added or used to replace items which are no longer desired. Although it is assumed that a longer questionnaire will have fewer returns than a short one, we have no pertinent evidence of this. Therefore, we should conduct a brief experimental study using two nonequivalent forms of the survey.

AD P000862

Organizational Maintenance Performance System

Michael Drillings, Ph.D.

US Army Research Institute

for the Behavioral and Social Sciences

In recent years, a change in Army training philosophy has resulted in a reduction in the length of initial school-based training for many Military Occupational Specialties (MOSs). Correspondingly, the responsibilities of units to provide on-the-job training (OJT) has increased. While the Army has increased the amount and variety of training material available to units, there is little evidence that the units have been able to use this material efficiently. The General Accounting Office (GAO) concurs in the finding that the management of training at the unit level is deficient.

When the Army Research Institute (ARI) first began to look at ways to increase the skills of Army maintainers, the lack of an overall effective training management system was seen as a hinderance to introducing effective OJT to units. While the recent introduction of the Battalion Training Management System (BTMS) appears to remedy some deficiencies, trainers of maintenance skills still lack the tools that specifically and easily point to the skills that most need training for each individual soldier.

ARI is now performing research to develop computer-based management systems for providing training and other information to managers. One system, for the direct support or intermediate level of maintenance, has been substantially completed and tested. Some of the training-relevant characteristics of this system are expected to be incorporated into the Standard Army Maintenance System (SAMS), which will be a computerized system for managing supply and maintenance operations. When this project is complete, it will be the first example of the inclusion of training information into a computerized Army system. The second system, for organizational maintenance, is still being developed in an armor battalion. The purpose of this report is to discuss the characteristics of that system.

Background

GAO (1978) has reported that Army maintainers are not receiving adequate on-the-job training at their units. They conclude that the reasons for these maintenance deficiencies are that unit commanders and supervisors are not sufficiently committed to develop OJT programs. The need for these programs should be clear. Dressel and Shields (1979), studying maintenance at the organizational level, found an average rate of unnecessary parts removal of 42%. For one item, relay boxes in the M551 turret, the unnecessary removal rate was 72%. Kern and Hayes (in press) also examined the performance of organizational maintainers. They found that on those tasks requiring an end-of-job checkout, 66% of the checkouts were either not performed or performed incorrectly. Furthermore, on tasks requiring special tools, 71% of the completed work contained uncorrected errors. Reports of deficiencies in maintenance performance are not limited to the Army. Orlansky and String (1981) report unnecessary parts removal rates for the other Services which are comparable to Army rates.

Why are soldiers having such difficulty maintaining their equipment? Before we consider our answer to this question, let us review some of the important characteristics of the Army training system at the unit level.

The current system of Army training places heavy emphasis on both the Soldiers' Manual and the Job Book. The Soldiers' Manual is a list of each task and its standard that soldiers in each MOS should be able to perform. The Soldiers' Manual, therefore, is a statement of overall training requirements. The Job Book, kept by each individual soldier's supervisor, is a record of each soldier's record of performance on each Soldiers' Manual task. The Job Book, therefore, should form the basis of an individualized training plan for each soldier. The problem with this system is that it is extremely rare for a Job Book to be kept accurately.

Although units have received the mission for conducting increased training that was once provided by the schools, the units have not received increased resources to accomplish that mission. Unit commanders are typically rewarded for having vehicles and equipment in good repair, not for training maintainers in how to do the repairs more effectively. When a commander must choose between high operational readiness (OR) rates and more training, he chooses high OR. Even when time is available for training, non-commissioned officers (NCOs), who must conduct the training, are generally not sufficiently trained to perform this job effectively. Also, even with BTMS, too many supervisors still view training as something that occurs behind a lecturn or in front of a blackboard and not on the shop floor.

Description of the Organizational Performance System

ARI and Anacapa Sciences, Inc. are now engaged in research to produce a new approach to the in-unit training of organizational maintenance skills.

The key elements of the approach are an overall model for unit OJT and a computerized information and evaluation system. Together, we refer to these elements as the Maintenance Performance System - Organizational (MPS-O).

In MPS-O, we have attempted to produce a unified system that will improve the current Army training system to more effectively train maintenance skills. One problem that we previously identified was the inaccuracy of Job Books. MPS-O tries to correct this deficiency by more clearly linking the performance of actual day-to-day operational maintenance and the record of that performance in the Job Book. Our electronic Job Book is kept automatically through inputs to the system that are primarily of operational and not of training value. Through an audit of the electronic Job Book, we know that its accuracy is substantially greater than that of the traditional Job Book.

In our experience, we have found that the electronic Job Book serves as the primary system output that controls training of maintenance tasks in the unit. MPS-O, however, contains a number of other outputs which also have the potential to motivate and direct training activities. The success of these outputs will be determined through further research.

Many of the other system outputs utilize the connection between individual maintainers and the vehicles upon which they work. This link permits supervisors to potentially use the vehicle repair history to identify both ineffective repairs and the individual responsible for the repair. These individuals can then be trained in the specific skills that they lack. As an added benefit, MPS-O maintains separate records of the number of preventive maintenance and corrective maintenance man-hours expended per vehicle. It also tracks the average number of man-hours expended for each specific task. This latter information can be used by managers to establish time standards for each task. Specific task repair times could then be compared to the standard with significant deviations possibly indicating a need for further training.

As part of the MPS-O model for the management and conduct of unit OJT, ARI has established a system for establishing and maintaining records on task qualification for each maintenance-related Soldiers' Manual task. The concept of task qualification has often been used informally in Army units over the years. Any system for qualifying maintainers on specific tasks seeks to take advantage of each individual's pride in his own ability and interest in receiving at least symbolic reward for a job well done.

In the MPS-O model, task qualification is awarded by each maintainer's supervisor according to easy to follow standards. In short, qualification indicates an individual's ability to perform the task correctly without supervisory intervention. The record of each maintainer's experience and qualification on each task is kept publicly in the shop area. We expect this to provide incentive both for maintainers who want to be trained and for supervisors by indicating that training is occurring.

One of the key problems previously identified is that commanders have few resources to provide for training activities and that NCOs often lack the ability to conduct formal training. For the MPS-O model, the solution to this problem is that training should occur, for the most part, as part of the normal operational maintenance responsibilities of the unit and be conducted by NCOs on the shop floor as part of their normal duties.

The structure of MPS-O is designed specifically to reinforce this concept of unit training. Maintainers are given credit for their performance of maintenance in the electronic Job Book and training requirements are identified by reference to how well normal maintenance activities are performed. Both maintainers and supervisors actively participate in the evaluation of system outputs. Maintainers can see how their record of experience and qualification changes as a function of supervised OJT and their supervisors can readily see their training activities reflected by those same changes. Commanders and higher level supervisors are also involved in the system because they receive summaries of both the training progress of individual companies and MOSs and overall reviews of the skills possessed by unit personnel.

While MPS-O is currently an experimental system, we are still planning for its future implementation into the Army. One major problem is that at this time, combat battalions do not have ready access to computers capable of supporting MPS-O. Fortunately, ARI is also conducting research with the high-technology division of Ft Lewis where computers are now being introduced at that level. Only after both units have demonstrated their ability to manage computer resources and the computers have demonstrated their ability to survive the battalion environment will MPS-O be capable of full implementation.

References

Dressel, J. D. and Shields, J. L., Organizational Maintenance Performance. ARI Research Problem Review, 79-8, 1979.

General Accounting Office, The Key to Improving Maintenance of Army Equipment: Commanders Must Motivate Their Personnel. LCD-78-428, 22 Dec 78.

Kern, R. P. and Hayes, J. F., Research Findings to Aid Supervisors and Trainers in Improving Maintenance Performance. ARI Technical Report, In Press.

Orlansky, J. and String, J., The Performance of Maintenance Technicians on the Job. IDA Paper. P-1597, Aug 1981.

AD P000863

A Comparison of Two Methods for Assessing Task Criticality

Eugene H. Drucker and R. Gene Hoffman
Human Resources Research Organization

David W. Bessemer
U. S. Army Research Institute

One of the problems experienced by training developers is the lack of sufficient time and resources to train all of the tasks that are performed by soldiers in a particular job or MOS. There are several ways to handle this problem. One is to lower the performance standards so that more tasks can be taught in the same amount of time. Another is to conduct training during hours that are not normally devoted to training, such as evening hours or weekends. Still another solution, and the one that is advocated by TRADOC in TRADOC Pamphlet 350-30, Interservice Procedures for Instructional Systems Development, is to limit training to those tasks that are most necessary for successful job performance. If the job is combat related, then training would be limited to tasks that are most necessary for the successful accomplishment of the unit mission. It is this concept, task criticality, with which this paper is concerned.

Guidelines for selecting tasks for training are also presented in TRADOC Pamphlet 350-30. Eight criteria are recommended, although the guidelines state that the user is free to use any of these criteria, and may even add new ones. A more recent set of guidelines for selecting tasks for training is contained in TRADOC Circular 351-4, Job and Task Analysis. This document states that the selection of critical tasks for training is one of, if not the most important requirement in the training developments area of responsibility. The document lists four criteria that should be used to select tasks for training:

1. The amount of delay that can be tolerated in performing the task.
2. The difficulty involved in learning the task.
3. The consequences of inadequate task performance.
4. The percentage of soldiers performing the task.

Based on these or other similar criteria, the training developer almost always obtains the required information from a questionnaire administered to a sample of job incumbents or experts. The respondents are given a list of task titles and are asked to choose a numerical rating according to each criterion for every task. The overall task criticality is then measured as a weighted average of mean ratings on each criterion. This method of collecting task criticality data has some inherent advantages for the training developer. It is easy to prepare, easy to administer, and easy to score. But the method has two potential deficiencies. First the criteria at best bear only an indirect relationship to mission accomplishment. Second, the criticality of a task may not be constant across all conditions. During some circumstances a task may be very critical, but during other circumstances it may not be critical at all. But what one actually finds in such surveys is that most tasks tend to be rated as being very critical. This may be due to the possibility that there is almost always some circumstance during which even a trivial task is important. In making their judgments on a task, respondents may tend to think mostly of situations where the task is important, without considering a variety of situations.

To overcome these problems, a different approach might be used. First, ratings of contribution to mission accomplishment can be obtained directly for each task. Secondly, particular situations in which the task is performed can be specified. A sample of combat situations would first be selected to include situations which are most likely to occur or situations for which we want soldiers best prepared. Once these situations are identified, a description or scenario could be prepared for each. Ratings of task criticality would then be based based on the role of each task within the scenario rather than on an abstract basis.

The purpose of this research was to compare judgments of task criticality based on the use of scenarios with judgments using the ISD approach. Three questions were asked. First, would the criticality of a task differ across scenarios? Second, would the criticality of a task obtained using the scenario-based method differ from those obtained using the ISD approach? And third, would the tasks selected for training differ when selected on the basis of these two approaches?

METHOD

To determine the effects of scenarios on judgments of task criticality and on the tasks selected for training, two surveys of task criticality were conducted at Fort Knox. The surveys differed in the type of questionnaire that was used. The questionnaire for one survey was based on the ISD guidelines and contained an alphabetical list of tasks that were to be rated along several ISD-based scales. The questionnaire for the other survey contained a set of scenarios. The tasks that were listed in this questionnaire were those that were described in the scenarios.

ISD-Based Questionnaire. The ISD-based questionnaire contained alphabetical lists of 161 tasks performed by platoon leaders of tank platoons during combat. These 161 tasks were selected from a larger number of platoon leader tasks identified in an analysis of armor operations conducted earlier by HumRRO for ARI. The 161 tasks were rated on four different scales. Three of the scales were prepared from the ISD guidelines contained in TRADOC Circular 351-4, Job and Task Analysis. The scales pertained to the amount of time required to learn the task by most new officers, the amount of damage to equipment and/or injury to personnel that could result from the performance of the task by the platoon leader, and the amount of time that the platoon leader would have available before starting the task. A fourth scale was prepared that would correspond more directly to the ISD definition of task criticality. This scale pertained to the effect of task performance on the successful accomplishment of the team mission. To eliminate the possibility that the order in which the scales were presented could affect the ratings of task criticality, the order of presentation was counterbalanced. Each scale contained five response alternatives from which the respondents were asked to select the most appropriate.

Scenario-Based Questionnaire. The scenario-based questionnaire contained four scenarios. Each described a different armor operation. The scenarios contained two parts. The first part was a description of the general situation. It described the team mission, the enemy situation, the terrain and weather, and the units involved in the mission or supporting it. A sketch depicted the situation described in the general mission. The second

part of the scenario depicted the special situation for the platoon. It described the role of the platoon in the mission and the tasks that were performed by the platoon leader. A second sketch depicted the special situation for the platoon.

Each scenario was followed by lists of the platoon leader tasks that were performed during the operation that was described. The tasks were presented in the same order in which they appeared in the scenario. The questionnaire contained a total of 51 platoon leader tasks. All 51 tasks were included among the 161 tasks contained in the ISD-based questionnaire. Several of the 51 tasks appeared in more than one scenario so that 66 ratings were made on each scale. To eliminate the possibility that the order in which the scenarios appeared could affect the ratings of task criticality, the scenarios were counterbalanced so that each scenario appeared in each of the four positions an equal number of times.

The tasks described in each scenario were rated on six different scales. Only one of these was identical to any of the four scales contained in the ISD-based questionnaire. This was the scale that pertained most directly to the ISD definition of task criticality. It asked the respondents to rate the effect of the performance of the task on the successful accomplishment of the team mission. The remaining five scales pertained to different factors that are generally regarded as affecting the outcome of combat. These were the effect of task performance on the effective application of fire power by the platoon; effective mobility and maneuver by the platoon; effective command, control, communication, and coordination, within the platoon; survivability of men and equipment within the platoon; and the capability of the platoon to sustain its combat effectiveness. Each scale contained five response alternatives.

Respondents. The respondents for both surveys were Army captains enrolled in the Armor Officer Advanced Course at Fort Knox. A total of 65 officers rated the 161 platoon leader tasks on the four scales based on the ISD method. A total of 57 officers rated the 51 tasks on the six scales contained in the questionnaire using scenarios. Each questionnaire was administered at the same time, but in different classrooms. The questionnaires were completed in approximately one hour.

RESULTS

The reliability of the scales contained in the ISD questionnaire ranged from .9? to .96, while the reliability of the six scales contained in the scenario-based questionnaire ranged from .86 to .93. Since only one rating scale appeared in both questionnaires, the remainder of this paper will be concerned only with it. This was the scale on which the respondents rated the effect of task performance on the successful accomplishment of the team mission. The reliability of the mission success scale was .94 in the ISD-based questionnaire and .89 in the scenario-based questionnaire.

Table 1 contains the five most critical and the five least critical tasks that were identified using the two types of questionnaires. The most critical tasks identified using the ISD method involved aspects of tactical decision making. Three of the tasks, Chooses Course of Action, Makes an Estimate of the Situation, and Analyzes Operations Order, involve either the

decision itself or the activities involved in the preparation for making the decision. The remaining two tasks, Issues Operations Orders and Issues Frag Order, involve the implementation of the tactical decisions. The most critical tasks identified in the scenario-based questionnaire are more combat specific and involve some aspect of gunnery. One task, Directs Enemy Be Engaged, appears twice. This is possible since the task appeared in more than one scenario.

Table 1

Tasks Rated As Most and Least Critical
Using ISD-Based and Scenario-Based Questionnaires

Most Critical	Tasks	Mean Rating
	1. Issues FRAGO	4.55
From ISD-Based Questionnaire	2. Issues FRAGO	4.52
	3. Chooses a course of action	4.49
	4. Makes an estimate of the situation	4.47
	5. Analyzes OPORD	4.45
	1. Directs fire and maneuver be conducted	4.48
From Scenario-Based Questionnaire	2. Directs enemy be engaged	4.46
	3. Directs avenue of approach be covered	4.45
	4. Directs enemy be engaged	4.39
	5. Requests indirect fires	4.39
Least Critical		
	1. Directs coil formation	2.69
From ISD-Based Questionnaire	2. Directs coil or herringbone formation	2.74
	3. Directs herringbone formation	2.78
	4. Controls interval between tanks	3.00
	5. Controls speed of tanks	3.05
	1. Requests illumination	2.84
From Scenario-Based Questionnaire	2. Monitors TOWs	3.50
	3. Directs ground guards be posted	3.59
	4. Requests wire communications be installed	3.61
	5. Directs air guards be posted	3.66

The least critical tasks identified using the ISD method involve aspects of tank platoon movement. Coil and herringbone formations are two formations used when a tank platoon halts along a route of movement. The remaining two tasks that were rated as least critical involved the interval between tanks and the speed of movement. The five tasks that were rated as being least critical using the scenario based questionnaire vary in their content.

One important aspect of these results should be noted. No task appears among the five most or least critical using both types of questionnaires.

Six of the 51 platoon leader tasks contained in the scenario-based questionnaire were embedded in more than one scenario or in more than one operation within the same scenario. One-way repeated measure analyses of variance were conducted to determine if the ratings of task criticality were

affected by the scenario in which they appeared. Table 2 lists the six tasks that were contained in more than one scenario and summarizes the results of the analyses. Significant effects for scenarios were obtained for two of the six tasks--Requests Indirect Fires and Submits SITREP. Thus, the results of the study showed that the context in which the tasks appear affected the ratings of task criticality for two of the six tasks appearing in more than one scenario.

The next step in the analysis was to compare the ratings of the 51 tasks that were contained in the scenario-based questionnaire with the ratings of the same tasks contained in the ISD-based questionnaire. Ratings were analyzed using a one-between and a one-within subjects repeated measures analysis of variance with method as the between subjects factor and tasks as the within-subjects factor. Four separate analyses were conducted, one for each scenario. If a task appeared more than once within the same scenario, the mean of its ratings was used in the analyses.

Table 2

Summary of Analyses of Variance Comparing
Task Criticality Ratings Obtained From Different Scenarios

Task	df	F	P
Directs Enemy Be Engaged	2,110	.97	ns
Requests Indirect Fires	4,216	2.68	.05
Requests Indirect Fires Be Adjusted	4,216	2.03	ns
Requests SPOTREPS	1,56	.25	ns
Submits SITREP	3,162	4.66	.01.
Submits SPOTREP	1,55	.21	ns

Table 3 summarizes the results that were obtained from the four analyses of variance. A significant effect for questionnaire type was obtained only in the analysis of the ratings obtained from the scenario Occupy Battle Position. When rated using the ISD-based questionnaire, the tasks contained in this scenario received an average rating of 3.72, while they received an average rating of 4.10 when rated using the scenario-based questionnaire. However, there was a significant task by questionnaire type interaction in all four analyses. These results indicated that the effects of questionnaire type varied with the task whose criticality was being measured.

Table 3

Summary of Analyses of Variance Comparing Task Criticality
Ratings Obtained From ISD-Based and Scenario-Based Questionnaires

	Main Effects For Questionnaire Type			Interactions Between Task and Questionnaire Type		
Scenario	df	F	P	df	F	P
Action on Contact	1,117	1.73	ns	1,1053	8.23	.01
Hasty Attack	1,116	1.14	ns	14,1624	4.17	.01
Occupy Battle Position	1,113	14.73	.01	22,2486	2.05	.01
Defend Battle Position	1,113	0.00	ns	10,1130	5.11	.01

Because of the significant interactions, post hoc comparisons were made for all tasks. Table 4 lists the number of tasks that appeared in each scenario and the number of tasks on which ratings of task criticality were affected by the method of measurement. The ratings of 2 to 4 tasks contained in each scenario were thus affected by the method of measuring task criticality.

The practical implications of these results can be illustrated best by examining the tasks that would be selected for training using each of the two types of questionnaires. For this analysis, only the 51 tasks that were contained in both questionnaires could be considered. Since the tasks that would be selected for training would also depend to a large extent on the number of tasks that could be trained, it was necessary to make an assumption about this number. The decision was arbitrarily made that about half or 25 of the 51 tasks could be selected for training. It was then necessary to decide how to handle the tasks that appeared in more than one scenario. The decision was made to use the highest criticality value received by a task if a significant difference was obtained for that task between scenarios. Otherwise, the mean rating across scenarios would be used. Of the 25 tasks that could be selected for training, 18 would have been selected by either of the two methods and 19 would have been rejected by either of the two methods. Seven of the 25 tasks selected for training would have been unique using either type of questionnaires. Thus, if half of the tasks could be selected for training, 28 percent of the tasks chosen for training would depend upon which method was used to assess task criticality.

Table 4

Number of Tasks On Which Ratings Of Task Criticality Were Affected By Type of Questionnaire

Scenario	Number Of Tasks	Number Significantly Affected By Questionnaire Type
Action on Contact	10	2
Hasty Attack	15	4
Defend Battle Position	11	2
Occupy Battle Position	23	3

One final observation should also be mentioned. The two surveys were administered simultaneously, but in different rooms. Neither group of respondents was aware of the purpose of the study, nor did they know that there were two different types of questionnaires. After the administration of the ISD-based questionnaire, several of the respondents in that group expressed their dissatisfaction with the questionnaire. They told the administrator that the survey was a waste of their time and that judgments of task criticality could not be made without providing some context in which to make the judgments. No critical comments were made in the adjacent room where the scenario-based questionnaire had been administered.

DISCUSSION

The results of this study have shown that the method used to assess task criticality affects the ratings of some tasks, but not all. From 13 to 27 percent of the tasks contained in the different scenarios received

significantly different ratings when different methods were used to assess their criticality. The results have also shown that the particular context in which a task is performed affects the criticality of some tasks, but not all. Two of the six tasks that were contained in more than one scenario received significantly different ratings when different scenarios were used to assess their criticality.

These results suggest that tasks differ in the degree to which their criticality is sensitive to the situations in which the tasks occur. This difference, unfortunately, can cause problems for the training developer who must select tasks for training, particularly when only relatively few tasks can be selected. If all of the tasks performed in a given job or MOS could be taught, then there would be no reason to even measure task criticality. But as increasingly fewer tasks can be taught, there is an increase in the potential impact of the particular method used to assess task criticality. When relatively few tasks can be chosen for training, there exists a possibility that the training developer could overlook a task that is critical in a high likelihood or high risk combat situation, either because the method used to assess criticality did not take combat situations into account or because it did not depict the particular situation.

The definition of task criticality on which this research was based specified that the criticality of a combat related task is equivalent to the degree to which the performance of the task affects the accomplishment of the unit mission. Ratings of the effects of task performance on mission accomplishment are only one way to measure task criticality. It can also be assessed, at least theoretically, by determining the actual effects of task performance on the outcome of different missions. But if a single rating is chosen to depict task criticality, then it is important that it be generalizable across combat situations. The results of this research have shown that not all ratings of task criticality are in fact generalizable. The criticality of some tasks must be assessed separately in the different situations during which the task is performed. The ISD method assumes the generalizability of all tasks, and its continued use to assess task criticality must therefore be reevaluated.

Finally, it is important to note some limitations to the present study. First, the two methods for assessing task criticality were compared for only a sample of leadership tasks performed by the platoon leader of a task platoon. It remains to be determined if similar results would be obtained for other types of tasks performed in combat. In addition, the effects of combat operations on ratings of task criticality were confounded with the effects of different combat situations since each operation was depicted as occurring in only one situation. The extent to which variations of task criticality are to be attributed to differences in combat operations as opposed to differences in combat situations remains to be clarified by future research.

AD P000864

The Jackknife: Its Application to Test Equating

James A. Earles
Malcolm James Ree
M. Gayle Kadura

Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks AFB, Texas

Many tests used by the Armed Services are revised frequently to update content and to reduce compromise. A major psychometric concern during revision is the necessity of deriving scores on the new test which are comparable to those on the old test. This score conversion permits the direct comparison of the scores of current examinees with those of past examinees and permits the retention of past decision score points because of consistency of meaning over time.

Equating Tests

The psychometric task of developing derived scores on one test so that they are equivalent to scores on a second test is called test equating or test calibrating. In general, two procedures have been in common use: linear equating and equipercentile equating. Linear equating equates two scores if their respective Z-score transformations are identical. It requires that equivalent Z-score transformations of the two tests represent the same cumulative proportion. Stated differently, distributional shapes must be the same or at least only trivially different. Equipercentile equating sets scores equal if they cut off equal proportions of the group or groups involved and makes no assumption of Z-score equivalence in the distributions. An equipercentile equating must be smoothed in some manner. The linear method offers the advantage of dealing with the analytic statistics of means and standard deviations, and because it provides a linear transformation, no smoothing is required.

Angoff (1971) has identified the two most frequently used equating designs as single group and equivalent groups. In the single group design, both the new test and the reference test are administered to the same group usually in a counterbalanced manner. In the equivalent groups design, each group is randomly given one of the two tests.

Jackknifing Equatings for Error Variance Estimates

The goodness of any given equating is difficult to assess. The stability of an equating depends upon sample size, the method, and the kind of smoothing performed. Error variance is one means of assessing the stability.

Variance error formulae for linear equating can be found in Angoff (1971). Lord (1981) gives variance error formulae for raw (unsmoothed) equipercentile equating. However, most equipercentile equating requires smoothing, and often extension to score points not found in the range of the equating sample. No variance error formulae exist for such smoothing or extension. There is a technique called "Jackknifing" (Miller, 1974) which can be applied in conjunction with any given equating method, giving results similar to equating in the usual manner and at the same time providing error variance estimates.

The technique known as "Jackknifing" is a method for obtaining an estimate on a sample and the variability of the estimate. It proceeds by estimating a sample statistic, Y, on the entire sample. The sample is then randomly partitioned into g independent non-overlapping groups, and $Y_{(i)}$ is the estimate in the total sample with the ith group removed. If g = 24, then 25 equatings are accomplished; 24 jackknife iterations for complements of the g groups and one equating for the total sample.

Let $Y_i = gY - (g - 1)\, Y_{(i)}$ be called a pseudo value and let

$$Y_* = \frac{1}{g} \sum_{i=1}^{g} Y_i$$ be the jackknifed value. The estimate

of the variance is given by setting

$$S_*^2 = \frac{1}{g}\left(\frac{\sum_{i=1}^{g} Y_i^2 - \frac{1}{g}\left(\sum_{i=1}^{g} Y_i\right)^2}{g-1}\right)$$

The Y_* and S_*^2 are then the estimate and its error variance respectively.

The jackknife may be applied to equipercentile equating which is smoothed and/or extended by any analytic technique. When the unjackknifed equating and its jackknifed version are similar, the jackknife technique will provide information on the stability of the equating at each test score value, including those values that are outside the range of the equating samples.

Problem

The purpose of this investigation was to compare the error variances given by jackknifing with the theoretical error variances from the formulae in Angoff (1971) for single and equivalent group linear equating. The object was to find a reasonable rule for the number of jackknife iterations to provide error variance estimates from jackknifing that are reasonable estimates of those obtained from the theoretical formulae. It is intended that this rule can be extended from linear equating to equating where there are no formulae for error variance (such as smoothed equipercentile equating).

II. METHOD

Samples

Applicant data on Armed Services Vocational Aptitude Battery (ASVAB) Form 8b and Form 9b were available. Data on 2,621 males for 8b and 2,587 males for 9b were collected at 21 Military Entrance Processing Stations and 800-plus associated outlying sites. Single group equatings were accomplished by equating different composites within ASVAB 8b, and equivalent group equatings were accomplished by equating composites from ASVAB 8b with composites from ASVAB 9b.

Composites

The ASVAB contains eight power subtests: Arithmetic Reasoning (AR-30 items), Auto-Shop (AS-25 items), Electronics Information (EI-20 items), General Science (GS-25 items), Paragraph Comprehension (PC-15 items), Mechanical Comprehension (MC-25 items), Mathematical Knowledge (MK-25 items), and Word Knowledge (WK-35 items). Four composites, based on these power subtests, were created for the equatings in this study.

	No. of Items	8b(9b) Skew	8b(9b) Kurtosis
EL = GS + AR + MK + EI	100	.23 (.24)	-.80(-.72)
MATH = AR + MK	55	.48	-.68
TECH = AS + MC + EI	70	-.25	-.69
VERB = GS + WK + PC	75	-.46 (-.18)	-.72 (-.93)

Equatings

For both single group equating (within the 8b sample) and equivalent group equating (between the 8b and 9b samples) the following were composites equated:

MATH (8b) to EL (8b) and MATH (8b) to EL (9b)-Similar distribution
MATH (8b) to VERB (8b) and MATH (8b) to VERB (9b)-Dissimilar distribution
TECH (8b) to VERB (8b) and TECH (8b) to VERB (9b)-Similar distribution

Analysis

The error variance formulae in Angoff (1971) for linear equating in single and equivalent group designs are both quadratic functions of the score values of the test being equated. It can be shown that for any given jackknifed linear equating, the error variance is also a quadratic function of the score values of the test being equated. The information from these quadratic functions is summarized as averaged standard errors where the average is taken over the range of possible score values on the equated test. The difference between the average from the formulae and average from the jackknifing was computed. Also, the ratio of the average from jackknifing over the average from the formulae was computed. This indicates how close the two are on the average and whether or not the jackknifed estimate is conservative.

III. RESULTS

The standard errors from both the formula and the jackknifed procedure were computed for the equatings at each value in the range of the equated test. The average difference over the test range between the formula and jackknifed standard errors is reported in Table 1 for jackknife iterations of size 5, 10, 25, 50, 75, 100, 150, and 250. Also reported in Table 1 is the ratio of the average jackknife error to the average formula error. Finally, information on computer (CPU) use is given for the various jackknife iteration sizes.

Table 1. Average Differences Between and Ratios of Jackknifed Standard Errors and Formula Standard Errors of Linear Equating

Number of Jackknife Iterations	Single Group Equating N = 2,624 Average Difference	Ratio	Equivalent Group Equating N = 2,587 Average Difference	Ratio
MATH to EL				
5	.081	1.42	.077	1.11
10	.039	1.12	-.031	.96
25	.031	1.16	-.010	.99
50	.026	1.13	-.024	.97
75	-.002	.99	-.012	.98
100	.003	1.02	-.013	.98
150	-.007	.96	-.037	.95
250	-.009	.95	-.083	.88
MATH to VERB				
5	.268	1.84	.239	1.42
10	.118	1.56	.051	1.09
25	.111	1.35	.040	1.07
50	.074	1.23	.036	1.07
75	.029	1.09	.026	1.05
100	.046	1.15	.013	1.02
150	-.077	.76	-.003	.99
250	-.003	.99	-.048	.92
TECH to VERB				
5	.470	2.50	-.081	.87
10	.280	1.89	-1.62	.73
25	.126	1.38	-.029	.95
50	.064	1.20	.032	1.05
75	.030	1.10	.056	1.09
100	.051	1.16	-.010	.98
150	.009	1.03	.012	1.02
250	.002	1.01	-.042	.93

Table 2. Average CPU Minutes for Various Jackknife Iterations

Number of Iterations	5	10	25	50	75	100	150	250
Average CPU Minutes	.19	.34	.79	1.55	2.30	3.05	4.56	7.58

IV. DISCUSSION

For single group equating, the results show that on the average the jackknifed estimates of error variance have a strong tendency to come down to and then go below the formula estimates as the number of jackknifed iterations increases. For equivalent group equating, such a tendency appears for the MATH to VERB equatings and somewhat for the MATH to EL equatings.

The objective was to find the point at which the number of jackknife iterations gives error variance similar to that given by the formulae. It is desirable to remain conservative in that the jackknifed averaged values should not be less than the averaged formula values. Fifty iterations appear reasonable. It should be noted that 50 is approximately the square root of the sample sizes. Further investigation on the generality of this jackknifing square root rule for providing variance estimates that are close but conservative with respect to the theoretical estimates would seem to be warranted.

Also, the computing time increases directly with the number of jackknife iterations. It should be noted that the absolute time is specific to this computer program. The jackknifing program reported actually produces four additional jackknifed equipercentile equatings in addition to the linear equating investigated.

In view of the similarity of jackknifed values to ordinary estimates and in view of the conservative nature of the jackknifed variance estimates, it would appear that use of the technique is advantageous in equating problems.

REFERENCES

Angoff, W. Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971, 508-600.

Lord, F.M. The standard error of equipercentile equating. RR-81-48, Princeton N.J.: Educational Testing Service, November 1981.

Miller, R.G. The jackknife-a review. Biometrika, 1974, 61, 1-15.

DISCIPLINARY PENALTIES AS SEEN BY SOLDIERS OF THE GERMAN FEDERAL ARMED FORCES

- Heinz Jürgen Ebenrett, Ph. D. -
- Jürgen Herrguth, M.A. -
Psychological Services of the German Federal Armed Forces

1. On the problem

The instrumentarium to maintain discipline and military order in the Federal Armed Forces is essentially graded in a twofold way. We distinguish between

- educational measures and
- disciplinary measures.

In the first instance both kinds of measures provide for a reward for exemplary performance of duty, for example by means of praise, award of a prize or pass. Related to the frequency of application, coercive measures, by means of which unwilling soldiers are disciplined, respectively disciplinary offenses are punished are, however, much more frequent. In the following we will confine ourselves exclusively to those coercive measures.

Educational measures are employed above all in order to encounter acute deficiencies in training and discipline. In most cases those measures are taken openly and directly. Depending on rank and office of the superior inter alia the following educational measures are possible:

A. general educational measures (every superior)
- correction, admonition, rebuke, warning
- order to remedy a deficiency
- prolongation of an exercise section
- report to a senior superior

B. additional educational measures (superiors: sergeants and higher ranks)
- physical exercises
- additional elaborations in writing
- additional repetitive duty (max. 1 hour)

C. special educational measures (disciplinary superiors only)
- change of the duty roster
- additional duty
- denial of overnight pass and weekend pass

If a culpable offense of duty, which is also important in kind and gravity was committed, as a rule a formal disciplinary penalty is indicated. Disciplinary penalties can only be imposed by the disciplinary superior, that is, by the company commander and higher ranks or - in particularly severe cases - by disciplinary courts. Disciplinary penalties include a formal examination and executive procedure and are reflected in the personal records of the soldiers concerned.

We distinguish between simple disciplinary penalties, which can be imposed by disciplinary superiors and disciplinary court measures:

Simple disciplinary penalties (disciplinary superior)
- reprimand
- strict reprimand
- disciplinary fine
- restriction of pass
- disciplinary arrest

Court disciplinary measures (disciplinary court)
- curtailment of salary
- ban on promotion
- reduction of rank
- dismissal

For the imposition of simple disciplinary penalties the "principle of expediancy" applies, that is, the responsible disciplinary superior determines at his own discretion, in conformity with his duty, whether and how steps must be taken against an offense of duty. Senior superiors may merely advise him, they may, however, not give any instructions. The relevant regulation (=Wehrdisziplinarordnung WDO (Military Disciplinary Code)) merely lays down that for the imposition the entire duty and off-duty conduct of the soldier concerned must be taken into consideration (Article 7 of the Military Disciplinary Code) and lays down that as a rule the more lenient disciplinary measures should be taken first and that only in the case of repeated offense of duty more severe disciplinary measures should be taken (Article 34 of the Military Disciplinary Code).

The two procedural regulations mentioned last emphasize the meaning of the Military Disciplinary Code which consists in providing the superior above all with an educational means to maintain discipline and order. The expiatory and retaliatory character of disciplinary penalties is clearly subordinated to this.

Since its introduction in 1957 the Military Disciplinary Code has basically remained unchanged. By the directive on "Educational measures" in 1970 it was, however, significantly supplemented.

Therefore the question arises whether the original function of the Military Grievance Code as an educational means is still valid today or whether it has been subjected to a change of meaning.

In the following examination steps we will investigate this question and by means of the frequency of the different disciplinary measures as well as on the basis of questioning data we will examine to what extent the educational meaning of disciplinary measures is still in the foreground of the imposition.

2. Basic data of the study

The present study forms a first step of analysis within the scope of a representative questioning by the Leadership Development and Civic Education Center at Koblenz, during the course of which in winter 1982/83 soldiers of all ranks are to be questioned on the effectiveness and

assessment of disciplinary measures. At present it is only based on the data of a preliminary questioning in summer 1982, by means of which the suitability of the questioning instrument was to be tested.

The random test of the preliminary questioning was composed as follows:

90 privates	Signal Battalion 310 at Koblenz
50 non-commissioned officers	Fighter Bomber Wing 33 at Büchel
52 disciplinary superiors	company commanders in the rank of a captain, who participated in a course of instruction at the Federal Armed Forces Command and General Staff College at Hamburg

Although the data may not be considered representative, they certainly permit the determination of valid tendencies, which can specifically be verified in the main study.

The data on the frequency of disciplinary measures which are also included in the analysis are, however, related to a respective overall counting of all soldiers in the Federal Armed Forces. The data are drawn from the statistical annual reports of the "Leadership and Civic Education Section" with the Armed Forces Office in Bonn.

3. Frequency of disciplinary measures

During the last ten years the number of simple disciplinary penalties decreased by more than 50%:

1972: 107,000 disciplinary penalties
1977: 66,000 disciplinary penalties
1981: 44,000 disciplinary penalties

The decrease in the number of disciplinary penalties may certainly not be attributed to a drastic improvement of discipline or supervision, although quite a number of the superiors questioned think that this is also an important reason (42.3% affirm an improvement of discipline; 34.6% an improvement of supervision). It must rather be assumed that since 1970 the instrument of educational measures is being employed to an increased degree instead of disciplinary measures also in the punishment of less severe offenses of duty.

Related to the career categories in 1981 proportions of disciplinary penalties were as follows:

Career category	percentage of disciplinary penalties	percentage of soldiers punished
privates	83.7%	9.6%
non-commissioned officers	15.5%	3.9%
officers	0.8%	0.6%

The frequencies of imposition reflect a clear imbalance among the career categories. According to that - related to the personnel strength of the career categories - privates are punished 16 times more often than officers and non-commissioned officers 6 1/2 times more often. Whether the different frequencies of penalties are merely a result of differences in the conduct of the personnel concerned or also a result of different handling of disciplinary power with regard to the members of different rank categories, must be left open. There are some facts in favor of the assumption mentioned last. Thus for example 75% of the superiors questioned admitted that under the same circumstances they inflict other disciplinary punishments on higher ranks than on lower ranks. Especially for officers an indirect, career-impeding effect would also be of great weight, which is expressedly excluded by the regulation on simple disciplinary penalties (Article 18.3 of the Military Disciplinary Code) and which 46.2% of the superiors questioned also regard as unjustified. Almost all disciplinary superiors (94.2%) are, however, convinced that even simple disciplinary penalties have a career-impeding effect and take that effect into account when taking disciplinary measures.

The differing treatment of the career categories also becomes obvious from the kinds of disciplinary penalties which are predominantly inflicted in the respective category.

Percentage of the simple disciplinary penalties in the career categories (1981)

	privates	NCO's	officers
reprimand	4.4	12.8	38.5
strict reprimand	6.4	21.7	25.1
fine	42.6	50.8	31.2
restriction of pass	31.7	7.1	-
arrest	14.9	7.6	5.2
	100.0%	100.0%	100.0%

On officers and non-commissioned officers almost exclusively less severe disciplinary penalties, on privates more severe disciplinary penalties are imposed. This drastic difference substantiates the assumption that obviously different direct and indirect effects for the members of different career categories are taken into account. That fact as well as the fact that on privates less severe disciplinary penalties are hardly imposed any longer is incompatible with the regulation (Article 34 Military Disciplinary Code). Furthermore it is opposed to the claim of the Military Disciplinary Code to have primarily an educational effect. Therefore in the following we will examine the question to what extent the actual handling of the disciplinary power is considered justified by the soldiers.

4. Attitude towards disciplinary measures

In accordance with the Military Disciplinary Code disciplinary superiors in the first line connect disciplinary penalties with the purpose

The ratios of approval make clear that the deterrent effect of disciplinary penalties - also as compared to other motives for a correct performance of duty - is estimated very highly by all career categories. Only the fear of discrimination in duty (e.g., extra duty, guard duty on weekends, cleaning of quarters and so on) is regarded as still more significant. It is, however, also remarkable that the majority of soldiers is also guided by intrinsic motives (sense of responsibility, acceptance of their roles) - with regard to the identification with their task as a soldier even to a much higher degree than estimated by the disciplinary superiors.

5. Educational claim of the disciplinary penalties

The vast majority of the soldiers are of the opinion that the educational effect of disciplinary penalties is overemphasized. By a great portion of all career categories it is even questioned altogether:

Ratios of approval with regard to opinions on education

- percentages of the answer alternative "correct" -

	privates	NCO's	officers
An adult cannot be educated any more	52.2	40.0	28.8
The educational effect of disciplinary penalties is overemphasized	84.3	58.0	48.0
Disciplinary penalties are an admission that education has failed	33.3	30.0	7.7
Disciplinary penalties above all result in obstinacy	73.3	70.0	

The answer alternatives confirm previous determinations according to which a primary educational effect of disciplinary penalties must be questioned. The opinion of more than two thirds of the privates and non-commissioned officers that disciplinary penalties above all result in obstinacy even speaks for a contrary effect. Obviously many soldiers consider the claim of their superiors who still want to educate them, inappropriate and out-of-date.

A further reason for the obvious negation of the educational effect of disciplinary penalties may be that the superiors do not meet the pedagogically necessary requirement which is also expressedly demanded in the regulation, that the special circumstances and the overall personality of the person to be rebuked must be taken into account. Thus for example only 36.5% of the superiors questioned were able to state that they always have enough time for their tasks as disciplinary superiors. The respective statements by the privates and non-commissioned officers still considerably support that assumption:

of bringing the rebuked person to reason. However, neither NCO's nor privates fully shared this optimistic attitude. The latter rather see the primary purpose of disciplinary penalties in a deterrence by threat of punishment. The purely punitive or retaliatory character of a disciplinary measure, however, is considered significant by none of the career categories:

primary function of disciplinary penalties

- percentage of the choice alternative "very important" -

	privates (90)	NCO's (50)	officers (52)
retaliation	10.0	4.0	-
deterrence of the rebuked person	25.6	20.0	13.5
deterrence of others	23.3	20.0	17.3
understanding of the rebuked person	17.8	22.0	69.2
understanding of others	21.1	28.0	34.6

It is remarkable that all groups consider the deterrent or educational effect of disciplinary measures on others more or less as important as the effect on the rebuked person himself. This finding is to a certain degree contrary to the regulation which expressively lays down that the imposition of a disciplinary measure must only be determined by the type of offense and the person concerned and which - except for a strict reprimand - forbids a publication of the disciplinary measure. It is, however, obvious that the general preventive function of disciplinary penalties is considered important by the vast majority of soldiers of all career categories.

In this connection the question is interesting how much importance is attached to the deterrent effect of disciplinary measures as compared to other motives for a correct performance of duty. For this we asked for the primary individual importance of different motives. The respective ratios of approval are as follows:

individual motives for a correct performance of duty

- ratios of approval in percent -

	privates	NCO's	officers*
personal sense of responsibility	56.6	88.0	67.3
acceptance of the role as a soldier	56.7	74.0	25.0
loss of prestige in the eyes of the comrades	37.8	44.0	21.1
fear of disciplinary penalties	72.2	54.0	65.4
fear of career disadvantages	37.8	68.0	45.1
fear of discrimination in duty	83.3	66.0	86.6

* officers: not their own motives but estimate of the motives of their soldiers

Assessments of the role of the disciplinary superior

-respective ratio of approval in percent -

	privates	NCO's
not enough time for subordinates	85.6	70.0
not enough knowledge on the private problems of subordinates	85.6	84.0
not sufficiently trained	45.6	32.0
too much power	47.7	24.0

Privates and non-commissioned officers agree that their disciplinary superior does not have enough time for them and does not know their personal problems. For that reason an individual educational claim can hardly be maintained. Nevertheless it is remarkable that the majority of soldiers do obviously not doubt the disciplinary power of their senior superiors, that is, they are not of the opinion that the superiors have too much power.

6. Summary

The vast majority of the soldiers consider disciplinary penalties very effective and regard the possibility of their infliction as justified. The importance as individual educational measure which is predominant according to the Military Disciplinary Code must, however, be questioned. On the one hand, because educational claims with regard to adults are obviously considered inappropriate to an increasing degree; on the other hand, because the disciplinary superiors do not have enough time to become sufficiently acquainted with their subordinates in personal respect as well. If at all, the individual educational influence is not exerted by the disciplinary superior, but by the direct superior (squad leader, platoon leader) by means of the instrumentarium of the "educational measures."

As opposed to that the central significance of disciplinary penalties seems to be a general deterrent effect. They are an effective risk treshold against offenses of duty and get things straight.

Military Eligibility and Participation in the All-Volunteer Force

Mark J. Eitelberg
Naval Postgraduate School

> Why, anybody can have a brain. That's a very mediocre commodity. Back where I come from we have great universities, seats of great learning where men go to become great thinkers. And when they come out they think deep thoughts, and with no more brains than you have. But, they have one thing you haven't got: a diploma.
>
> --The Wizard to the Scarecrow in *The Wizard of Oz* (1939)

The Wonderful Wizard was truly a wiz if ever a wiz there was. Everyone has a brain. Some may even have the capacity to think great thoughts. But, in the final analysis, people are just folks, and it doesn't matter a hoot whether your head is stuffed with grey matter or little bundles of straw. The main mark of distinction is the educational equivalent of a red badge of courage: pieces of paper with foreign words, lots of loops and curls, gold seals, and impressive signatures.

In some ways, the leaders of this country's modern military share a perspective not unlike that of the Great and Powerful Oz--and the similarities even extend beyond a mutual attachment to the color green. For, in the world of the military's policymakers and data analysts, in the realm of placement officers and recruiters alike, diplomas and degrees hold an almost mystical property. With diploma in hand, accompanied by a reasonably high score on the standardized entry test, the fabled strawman himself could enlist in any one of the Armed Services with favorable opportunities for technical training, special benefits, and career advancement. Moreover, because the amiable Scarecrow is a bonafide recipient of the treasured document, he stands a much better than average chance of fulfilling his initial term of enlistment in praiseworthy fashion.[1]

Measures of "Quality" and Eligibility for Military Service

"Quality," in the Department of Defense lexicon, generally refers to those characteristics and attributes of military personnel that are considered desirable and that contribute to a more productive, better motivated, and highly capable force. Because of the difficulty in constructing individual profiles and deriving measures of motivation and performance--and because of the wide range of different occupations in the Armed Services--manpower "quality" is customarily described in the shorthand terms of educational level and standardized test scores.

The Armed Services place a high premium on completion of high school.[2] It is commonly accepted that "possession of a high school diploma is the best single measure of a person's potential for adapting to life in the military."[3] Male enlistees who have not completed high school (at time of entry), for example, are about twice as likely as are high school graduates to leave the military before finishing their full first term of active duty. In addition, non-high school graduates typically experience more disciplinary, administrative, and retraining actions.[4] Consequently, "the active force recruiting programs have concentrated on enlisting high school diploma graduates."[5] The practical gauge of military recruiting "success" since the end of conscription in December 1972 has been the comparable proportion of high school graduates in the general population--even though the Military Services attempt to recruit as many high school graduates as possible in any given year through the use of quotas, enlistment bonuses and other special incentives, and differential qualifying standards.

As in the case of formal education, the Services would prefer to recruit the "best and the brightest" young men and women from the general population. The experience of the last thirty-five years suggests that individuals who score relatively low on the military's aptitude test tend to be less successful in training programs than those who score in the higher range. In addition, evidence shows that higher-scoring recruits are less likely to have disciplinary problems and more likely to

---

[1]The Cowardly Lion, if so inclined, could serve his country quite effectively along with Toto in the Canine Corps. The Tin Woodman, because of his steely nature, might very well be eligible to serve in one of the Army's Infantry/Armor specialties. And dear Dorothy, of course, could remain close to her home and Aunty Em by signing on with the Kansas National Guard.

[2]Officers are normally required to have a college degree. The issue of educational quality in the AVF is therefore focused primarily on the enlisted ranks.

[3]Department of Defense, America's Volunteers (Washington, D.C.: Office of the Assistant Secretary of Defense [Manpower, Reserve Affairs, and Logistics], December 1978), p. 30.

[4]Department of Defense, Defense Manpower Quality Requirements, Report to the Committee on Armed Services of the U.S. Senate (Washington, D.C.: Office of the Assistant Secretary of Defense [Manpower and Reserve Affairs], January 1974); and General Accounting Office, Problems Resulting from Management Practices in Recruiting, Training, and Using Non-High School Graduates and Mental Category IV Personnel (FPCD-76-24) (Washington, D.C.: General Accounting Office, 12 January 1976).

[5]Department of Defense, America's Volunteers, p. 30.

develop the requisite skills to be effective on the job. "Though there are many high-scoring personnel who prove ineffective and many low-scoring persons who perform well," the Department of Defense points out, "on the average, the higher an individual's [aptitude test] score, the greater the likelihood of successful military performance."[6]

The test used to screen applicants for enlistment is the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB consists of ten subtests. The scores of four of the subtests (Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Numerical Operations) are combined to produce an Armed Forces Qualification Test (AFQT) score. The AFQT score, supplemented by scores on various composites of aptitude subtests, is used in conjunction with educational, medical, and moral standards to determine an applicant's enlistment eligibility. Scores on aptitude composites are also used to determine an applicant's eligibility to enter training in specific military occupations.

...nt Eligibility and Participation in the Volunteer Military: A Portrait of Contemporary Youth

... 1980, the Department of Defense and the Military Services, in cooperation with the Department ...or, sponsored a large-scale research project to assess the vocational aptitudes of American ...th. A national probability sample of approximately 12,000 young men and women, selected from participants in the National Longitudinal Survey (NLS) of Youth Labor Force Behavior, was administered the ASVAB.

This major research endeavor, known as the "Profile of American Youth," marks the first time that a vocational aptitude test has been given to a nationally representative sample. The "Profile" study thus offers an unpredecented opportunity to evaluate the "cross-sectional character" of military enlistees based on a national measure of vocational test performance.

The "Profile" study sample contains approximately equal proportions of males and females, including individuals from urban and rural areas, and from all major census regions. For the purposes of previous analyses, this sample was statistically weighted to correspond with the 1980 national youth population. Since the "Profile" study incorporates the scores of contemporary youth on a similar version of the ASVAB used currently to screen military recruits, it is possible to estimate, with reasonable precision, the numbers and proportions of American youth who would be expected to qualify for military enlistment under present standards. Enlistment eligibility rates for the general population, when combined with information on enlistment behavior, also allow--for the first time--accurate computation of the military "participation rates" of qualified youth.

Numerous attempts have been made throughout the years to fix the limits of the so-called "eligible" population and, therefrom, to calculate the military "participation rates" of various demographic subgroups.[7] The rates of participation for all youth (or specific age cohorts) can be easily determined with Department of Defense statistics (Master/Loss data files) and Bureau of the Census population estimates. However, the "participation rates" of qualified youth--a more "refined" measure of participation--must be based on a reasonable estimation of the number and characteristics of potentially qualified youth. Most attempts to describe the pool of potentially qualified youth have, in the past, hinged upon aptitude test score data compiled for pre-inductees or the aggregate population of applicants/examinees. Consequently, previous estimates of the "participation rates" of potentially qualified youth are subject to serious error.

Each Military Service applies its own aptitude standards in determining eligibility for enlistment. These aptitude standards reflect the diverse requirements of the separate Services, and they typically vary according to educational attainment (high school graduation status) and, at times, according to sex. For example, in the Army, male and female high school graduates during FY 1981 were required to achieve a minimum AFQT score of 16 and a score of at least 85 on one of nine Service-specific aptitude composites. In contrast, Air Force enlistment standards for FY 1981 required that male and female high school graduates achieve a minimum AFQT score of 21; in addition, they were required to attain a combined aptitude composite score (including the Mechanical, Administrative, General and Electronics composites) of no less than 120.

---

[6]Department of Defense, Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery (Washington, D.C.: Office of the Assistant Secretary of Defense [Manpower, Reserve Affairs, and Logistics], March 1982), p. 7.

[7]Examples of previous research include: R.V.L. Cooper, Military Manpower and the All-Volunteer Force (R-1450-ARPA) (Santa Monica, CA: Rand Corporation, 1977), pp. 213-216; B.D. Karpinos, Qualification of American Youths for Military Service (Washington, D.C.: Office of the Surgeon General, Department of the Army, 1962), and several other publications by the same author; C. Kim et al., The All-Volunteer Force: An Analysis of Youth Participation, Attrition, and Reenlistment (Columbus, OH.: Center for Human Resource Research, Ohio State University, May 1980); and Directorate for Manpower Research, Geographic and Racial Differences Among Men Qualified for Military Service (Research Note 72-16) (Washington, D.C.: Office of the Assistant Secretary of Defense for Manpower and Reserve Affairs, July 1972) and subsequent reports by the Manpower Research and Data Analysis Center. The other side of the issue--the characteristics of the population considered unqualified for military service--is treated in The President's Task Force on Manpower Conservation, One-Third of a Nation: A Report on Young Men Found Unqualified for Military Service (Washington, D.C.: Government Printing Office, January 1964).

Higher aptitude scores are required ordinarily for male non-high school graduates and recipients of General Educational Development (GED) high school equivalency certificates in each of the Services. In FY 1981, female non-high school graduates were not eligible for enlistment in either the Navy or the Marine Corps; and female high school graduates who wished to enlist in these Services were required to meet different aptitude standards than those established for males.

Recent analyses by the Human Resources Research Organization (HumRRO) and the Brookings Institution--using the separate Service aptitude standards in effect during FY 1981--have been performed to determine (on the basis of ASVAB results and data on sex and education) the numbers and proportions of American youth (ages 18 through 23) who would qualify for military service.[8] Aptitude standards for FY 1981 were used because this period (October 1980 through September 1981) coincides roughly with the point of educational attainment established for the "Profile of American Youth" population (i.e., September 1980, or the start of the 1980-81 school year).

Table 1 displays the results of the HumRRO and Brookings analyses. First of all, it is apparent that enlistment "selectivity" varies from Service to Service. Proportionately more American youth, regardless of sex, would be expected to qualify for the Army than for any other Service. At the same time, the lowest proportion of youth would be expected to qualify for the Marine Corps. The stringent Marine Corps "selectivity quotient" is largely the effect of entry restrictions on females. The Navy's debarment of female non-high school graduates also affects the eligibility rate for all youth in this Service. Not shown in Table 1 are the separate eligibility rates for males and females. The estimated eligibility rates for all male youth, by Service, are as follows: Army, 77 percent; Navy, 75 percent; Marine Corps, 72 percent; and Air Force, 63 percent. The estimated eligibility rates for all females are: Army, 80 percent; Navy, 58 percent; Marine Corps, 46 percent; and Air Force, 60 percent.

Table 1

Estimated Percent of American Youth (18-23 Years) Who Would Qualify for Enlistment in the Military Services By Racial/Ethnic Group and Educational Level[a]

Racial/Ethnic Group and Education[b]	Military Service: Army	Navy	Marine Corps	Air Force
White[c]				
NHSG	41.7	19.9	22.5	11.2
GED	76.0	70.4	35.1	56.1
HSG	96.4	87.5	79.8	85.1
TOTAL	85.7	74.5	67.7	70.5
Black[d]				
NHSG	7.1	[illegible]	3.9	0.8
GED	35.2	26.6	13.9	11.2
HSG	68.6	45.6	33.6	32.1
TOTAL	[illegible]	31.7	23.6	21.5
Hispanic				
NHSG	13.6	4.8	5.5	1.5
GED	40.0	35.7	18.8	16.8
HSG	65.7	64.8	54.7	56.7
TOTAL	54.6	39.2	33.3	32.7
TOTAL				
NHSG	31.6	15.0	16.8	8.0
GED	68.0	62.1	31.1	47.4
HSG	92.7	81.6	73.2	77.6
TOTAL	78.7	66.6	59.6	61.5

Source: M. Binkin and M.J. Eitelberg with A.J. Schexnider and M.M. Smith, Blacks and the Military (Washington, D.C.: The Brookings Institution, 1982), p. 38; and special tabulations provided by the Office of the Assistant Secretary of Defense for Manpower, Reserve Affairs, and Logistics.

[a]Estimates of the percent of youth qualified for military service were calculated on the basis of results from the Profile of American Youth (administration of the Armed Services Vocational Aptitude Battery [ASVAB] to a national probability sample in 1980) and the 1981 education/aptitude standards used by the Armed Services. (It should be noted that eligibility for enlistment would also depend on other factors--including medical and moral requirements.)

[b]NHSG is non-high school graduate. GED is recipient of General Educational Development (GED) high school equivalency certificate. HSG is high school diploma graduate or above. The American youth population includes all persons born between January 1, 1957 and December 31, 1962. Educational level was determined as of September 1980 (start of 1980-81 school year).

[c]White category includes all racial/ethnic groups other than Black or Hispanic.

[d]Black category does not include persons of Hispanic origin.

[8]See Martin Binkin and Mark J. Eitelberg with Alvin J. Schexnider and Marvin M. Smith, Blacks and the Military (Washington, D.C.: The Brookings Institution, 1982).

The differences in the enlistment eligibility rates for the three racial/ethnic groups displayed in Table 1 are quite substantial. For example, approximately four out of five white youth would be expected to qualify for enlistment in the Army. Just over half of all Hispanic youth, and just under half of all black youth, would meet the minimum aptitude standards established by the Army. And the disparity between racial/ethnic groups is even wider in the other Services. About three out of ten white youth, for instance, would probably fail to qualify for entry into the Air Force, based on FY 1981 minimum aptitude/education standards; in sharp contrast, almost four out of five black youth would probably be rejected by the Air Force.

Substantial variance in the eligibility rates of youth by educational level can also be observed both within and between separate racial/ethnic groups. The enlistment eligibility rates for non-high school graduates, regardless of racial/ethnic group, are considerably below the comparable rates for persons with equivalency certificates or high school diplomas. Minorities who are high school dropouts (without GED certificates), in fact, have little or no likelihood of being able to meet the minimum enlistment criteria established by the Armed Services.

Table 2 displays the estimated numbers of young men and women (totals by racial/ethnic group and Service only) who would be expected to qualify for enlistment. These data give some idea of the approximate number of youth affected by the eligibility rates shown above--as well as the differential impact of Service standards on the supply of qualified applicants. (A forthcoming report by HumRRO will present the percentages and numbers of American youth who would be expected to qualify for military service--according to racial/ethnic group, educational level, gender, and geographic region--under the same standards outlined here.)

Table 2

Estimated Number of American Youth (18-23 Years) in the General Population and the Estimated Number Who Would Qualify for Enlistment in the Military Services by Racial/Ethnic Group [a]

(Number in Millions)

Racial/Ethnic Group[b]	Number in general population	Number Qualified for Military Service			
		Army	Navy	Marine Corps	Air Force
White	20.1	17.2	15.0	13.6	14.2
Black	3.4	1.6	1.1	0.8	0.7
Hispanic	1.5	0.8	0.6	0.5	0.5
TOTAL	25.1	19.6	16.7	14.9	15.4

Source: Derived from special tabulations provided by the Office of the Secretary of Defense (Manpower, Reserve Affairs, and Logistics).

[a]Base population includes residents of the United States born between January 1, 1957 and December 31, 1962. Base population figures in this table exclude persons for whom education was unknown. Exclusion of these persons reduced base population figures by an average of 1.4 percent below Bureau of the Census estimates. Unknown cases occurred most often among black males (2.2 percent) and least often among Hispanic and white males (1.2 percent).

[b]White category includes all racial/ethnic groups other than black or Hispanic. Black category does not include Hispanic.

The military "participation rates" of American youth (males only) were calculated with data from the "Profile of American Youth" study and recruiting statistics compiled by the Defense Manpower Data Center. The "participation rate" is defined as the percentage of male youth born between January 1, 1957 and December 31, 1962 who enlisted in the military (for the first time) between July 1973 and September 1981.

Table 3 shows the participation rates, by racial/ethnic group and educational level, for two base populations: (1) all male youth (within the respective category); and (2) all male youth who would be expected to qualify for enlistment under FY 1981 aptitude test standards (by racial/ethnic group and education category). It should be noted that the cross-sectional participation rates displayed in Table 3 actually understate the true percentages of male youth who join the military, since they do not include individuals who either (a) enlist after September 30, 1981 or (b) enter officer programs. It should also be pointed out that eligibility for enlistment would depend on other factors in addition to aptitude and education--including medical and moral requirements.

Table 3

Military Participation Rates of Male Youth Born Between 1957 through 1962 by Racial/Ethnic Group and Educational Level[a]

Educational Level[b]	Racial/Ethnic Group White[c]	Black[d]	Hispanic	TOTAL
Below High School Graduate				
All Youth	16.6	12.1	5.3	14.5
Qualified Youth	39.0	135.7[e]	45.7	45.1
GED High School Equivalency				
All Yuuth	18.6	14.2	14.5	18.0
Qualified Youth	25.5	37.6	29.7	27.0
High School Diploma Graduate and Above				
All Youth	9.8	22.3	10.3	11.2
Qualified Youth	10.2	33.7	11.6	12.2
TOTAL				
All Youth	11.5	18.2	8.3	12.3
Qualified Youth	13.6	41.6	15.3	16.0

Sources: Statistics on qualified youth are derived from data that appear in Department of Defense, Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery (Washington, D.C.: Office of the Assistant Secretary of Defense for Manpower, Reserve Affairs, and Logistics, March 1982); and special tabulations provided by the Office of the Secretary of Defense.

[a]Participation rate is the percentage of male youth born between January 1, 1957 and December 31, 1962 who enlisted in the military (for the first time) between July 1973 and September 1981. Participation rates are shown for two base populations: 1. all male youth within the racial/ethnic and education category; and 2. all male youth who would be expected to qualify for enlistment under 1981 aptitude test standards (by racial/ethnic and education category). The cross-sectional participation rates understate the true percentage of male youth who join the military since they do not include individuals who a) enlist after 30 September 1981 and b) enter officer programs. Estimates of the number of youth qualified for the military were calculated on the basis of results from the Profile of American Youth (administration of the Armed Services Vocational Aptitude Battery to a national probability sample in 1980) and the 1981 education/aptitude standards used by the Armed Services. (It should be noted that eligibility for enlistment would also depend on other factors--including medical and moral requirements.)

[b]For military personnel, education at time of entry (and initial qualification) into service. Approximately one percent of the male youth population could not be identified on the basis of education; and one percent of military personnel could not be identified on the basis of racial/ethnic group. These unknown cases were not included in the calculations of participation rates.

[c]White category includes all racial/ethnic groups other than black or Hispanic.

[d]Black category does not include persons of Hispanic origin.

[e]During FY 1976-80, the Armed Services unknowingly accepted volunteers who did not meet eligibility standards because of errors in test calibration. These errors affected principally non-high school graduates with low aptitude scores. The unusually high "participation rate" for black non-high school graduates reflects the fact that many more black youth in this category were accepted for military service than would have qualified with the correctly calibrated test.

The attraction of the military for minority youth is vividly portrayed in Table 3. Black and Hispanic youth who are qualified for military service have generally enlisted in proportionately greater levels than their white counterparts. This is particularly true for blacks: as of September 1981, almost 42 percent of all potentially qualified black males in the United States (born in 1957 through 1962) have entered military service. One out of three black male youth who had a high school diploma or a GED, and would probably qualify for enlistment, had enlisted by September 1981--while the comparable rate for black high school dropouts is a whopping 136 percent. (This unusually high rate reflects the fact that ASVAB misnorming during FY 1976-80 affected principally the eligibility of non-high school graduates with low aptitude test scores. Many more black youth in this category consequently were accepted for military service than would have qualified with the correctly calibrated test.) In contrast, the participation rate for potentially qualified white high school graduates is 10 percent; and the overall rate for white males who would qualify for enlistment is about 14 percent.

Perhaps an even more revealing aspect of youth participation lies in the fact that potentially qualified youth who do not have a high school diploma or equivalency certificate--regardless of race--find military service an especially appealing job or education alternative. Almost half of all high school dropouts who could probably pass the military's aptitude test standards had enlisted; and more than one out of four qualified GED recipients had made the same choice. In fact, the impact of the Armed Services as a place of opportunity, equal acceptance and involvement, regardless of prior social disadvantage or pre-existing handicap, has helped to make the military a traditional channel for social mobility. The participation rates displayed in Table 3 tend to confirm that both the image and the promise of "opportunity" are still quite strong.

Some General Observations

As a matter of fact, our fantastic friends from the Wizard of Oz may pass the military's education/aptitude requirements. Their perseverence in getting to the Emerald City and the Scarecrow's diploma make them good risks insofar as the completion of their first term of duty. With "passing" scores on the AFQT, they would be eligible to join the enlisted ranks. It is highly questionable, however, whether Dorothy's three strange companions could ever meet the medical standards established for military eligibility. (And, alas, the poor Scarecrow himself would surely be a fire hazard.)

In the real world, nevertheless, the Military Services are faced with the task of selecting--from among almost a million potential recruits each year--hundreds of thousands of the nation's very "best" prospects. And for several hundred thousand young men and women annually, acceptance or rejection by the Armed Forces will affect not only their immediate opportunities for employment and training, but the total sum of their early "life chances" and the eventual course of their working life. For some young men and women, service in the nation's military may even be a sort of crossroad or junction between a path to socioeconomic "failure" or "success."

Recognition of the consequences of personnel screening decisions in the Armed Forces--on the individual "life chances" of today's youth as well as the nation's own defense capabilities--has operated to place the military's enlistment criteria under greater scrutiny than ever before. As the authors of one recent study observe: "Whether the standards used for enlistment, job classification, and assignment are as valid as adherence to them implies is an open question. While in many cases present standards are based on years of experience and are the products of extensive and rigorous research, in others they appear to be nothing more than legacies of the conscription era when there was virtually no pressure on the armed forces to justify their manning criteria."[9]

Congress has strongly urged the Department of Defense and the Military Services to develop an empirical research and analytical foundation for enlistment standards presently in use.[10] Indeed, major efforts are currently underway to validate existing standards and to expand the selection and classification measures applied by the military (particularly aptitude test scores). Research is also in progress now to include consideration of various high school credentials, additional aptitude test scores, high school academic records, and attendance and behavioral records in an effort to refine further the recruit screening process. For example, it has been noted that, with the wide and almost limitless variety of high school "graduation" standards being used in the various states, school districts, and individual secondary schools, the current educational standards applied by the Armed Forces appear almost arbitrary. More "precise" standards, it is felt, can be developed to coincide with the substantial changes that have occurred in the secondary school systems of this country over the past two decades. Clearly, some applicants who should not be allowed to enlist are being accepted; on the other hand, it is quite possible that many individuals who would probably perform well in the military are being eliminated from consideration due to educational standards that are outdated, unnecessarily rigid, imprecise, and overly generalized. Current and future research efforts--including testing research, an assessment of educational and moral standards, a reexamination of medical criteria, and the ongoing analysis of the "Profile of American Youth" data base--should help the scientific and policymaking community evaluate the standards presently used by the Armed Forces as the basis for their personnel decisions--and, at the same time, reach a more complete understanding of the relationship and role of the military in society.

---

[9]Binkin and Eitelberg, Blacks and the Military, p. 155.

[10]Department of Defense, Department of Defense Efforts to Develop Quality Standards for Enlistment, Report to the House and Senate Committees on Armed Services (Washington, D.C.: Office of the Assistant Secretary of Defense [Manpower, Reserve Affairs, and Logistics], December 1981), p. 1.

ABSTRACT

Criterion-Referenced Testing for Technical Training*

John A. Ellis, PhD

Navy Personnel Research and Development Center

Modern military instruction is developed according to a systematic method called Instructional Systems Development (ISD). Testing of student performance is an important part of ISD, since the adequacy and maintenance of any training program depends on careful assessment of the quality of student learning. To facilitate the test and test items development process, the Handbook for Testing in Technical Training was developed. This paper described the handbook and discussed the role of criterion-referenced testing in technical training.

* NOTE: Material contained in this presentation will be included in a forthcoming book to be published under the auspices of the Military Testing Association and tentatively entitled Military Contributions to Development of Training Technology. (Editors: John A. Ellis, Navy Personnel Research and Development Center, and Hendrick W. Ruck, Air Force Human Resources Laboratory.)

AD P000867

INSTRUCTOR ATTITUDES TOWARD INSTRUCTIONAL SYSTEMS DEVELOPMENT AND PERFORMANCE BASED TESTING

By: Homer C. Emery, Maj, MSC, USA
Florence P. Emery, Education Specialist, AHS
Ben Pierce, Education Specialist, AHS

This paper focuses on an issue often talked about among professional training developers and frequently discussed in the literature. (1,2,3). The issue which we address is instructor attitudes toward instructional systems development and performance based testing.

First let me admit that I am an instructor, or as normally referred to, "the subject matter expert (SME)". My first experience with ISD was at Fort Sill, Oklahoma, where one of my colleagues, Ms. Florence Emery, was involved in instructional development for subjects ranging from "crater analysis" to "map reading". My first attitudes toward ISD were formed on the question of "How can a non-subject-matter expert develop effective technical training better than an SME?" At that time, for me, this question was an honest concern.

This same concern is very real for many trainers in organizations that are implementing an instructional development system that involves professional training developers rather than subject matter experts. When this concern is ignored, neglected, or simply overlooked it may have negative impacts on delivery of the final training product.

During discussions with the co-authors it appeared that identifying instructor attitudes toward ISD and performance based training would be helpful in the implementation of new training programs. We decided to develop an instrument that could be used for assessing instructor attitudes that could impact on training. Our original project was designed in three phases:

Phase one: Develop statements for use in an attitude assessment instrument.
Phase two: Conduct formal item analysis of the assessment instrument.
Phase three: Apply the instrument in an actual training environment.

In this paper we will be reporting on what we have accomplished in Phase one of this project and discuss preliminary observations concerning instructor attitudes toward ISD and performance based testing. In addition several strategies for implementing ISD in traditional training organizations will be presented.

Before we continue it will be helpful to first define attitude and develop an operational definition of "instructor attitudes". From Lang's early description of "aufgab" (task-attitude) the concept and definition of attitude has been argued and debated. Some experts have even suggested that the concept of attitude should be abandoned.

A dictionary would define attitude as: "The mental posture or position in relation to some purpose or emotion." Numerous definitions of attitude can be found in the literature. Thurstone provided an early definition that we will share with you, "An attitude is the sum total of a man's inclinations

and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any specified topic."(4)

For the purposes of this paper instructor attitudes are defined as, "concerns of the instructor about ISD, performance based testing, and the training environment that may influence the training process."

The first problem encountered was the selection and development of an instrument for assessing instructor attitudes. Several methods and instruments for ascertaining attitudes have been successfully used by other workers. It was decided to use the Likert Summated-Rating method due to its relative simplicity and the time required for construction. Developed by Rensis Likert in 1932 the method of summated ratings consists of a series of statements to which an individual may respond by indicating a range of concurrence. Each statement represents different aspects of the attitude object.

Thirty-nine statements were constructed from training journals and training development publications. These statements represented the following areas of possible instructor concern:

*Concerns about testing.
*Concerns about Instructional Systems Development.
*Concerns about student abilities.
*Concerns about training methods and design.

Responses from subjects were obtained by use of five categories: strongly agree (SA); agree (A); undecided (UN); disagree (D); and strongly disagree (SD). Numerical scores of 5-1 were used to indicate the degree of response as possitive or negative.

To minimize possible set response from test subjects approximately half the statements were designed so that scoring was reverse. This is illustrated in the following statements:

Positive Statement -
Students should be provided with an outline of the steps of a performance test prior to testing.
SA(5); A(4); UN(3); D(2); SD(1)

Negative Statement -
A few students can always be expected to fail a course.
SA(1); A (2); UN(3); D(4); SD(5)

The total attitude score for an individual is obtained by summing the numerical scores of individual statements.

To determine which of the original statements would be selected for formal item analysis a preliminary investigation was conducted utilizing the criterion of internal consistency. In this method the summed response from 25% of the highest scoring individuals is compared to the summed responses from 25% of the lowest scoring individuals. The difference in summated ratings is used to rank the statements.(5)

In our preliminary analysis the original instrument was given to a group of training development professionals (n=40). This group was selected from the membership of professional training development organizations and was assumed to be a known positive group (KPG). The instrument was also given to a group of military instructors (n=40) associated with the 91S MOS (environmental health specialist) course at the Academy of Health Sciences, Fort Sam Houston. The instructor group represented an unknown group (UNG).

In this presentation only selected statements will be presented for discussion. Responses to the following statements indicated a difference in how developers and instructors viewed testing. "A test that all students can pass is too easy and should be modified." The summated rating for the known positive group (KPG) was 41 and the unknown group (UNG) summated rating was 26. "Some test items need to be more difficult to enable ranking of students." The summated rating for the KPG was 42 and the UNG 27. The instructor group tended to view testing as a means of numerical ranking of students rather than as an indicator of performance ability.

The second area in which statements were constructed was that of concern about ISD. Response to the following statements indicage an uncertainty on the part of instructors about ISD. "ISD is less effective than traditional methods." The summated rating for KPG was 42 and the UNG 26. "ISD developed training weakens the instructor's role in the classroom." The summated rating for the KPG was 41 and the UNG 19. This may indicate that instructors may feel threatened about the ISD process.

The third area addressed was that of student abilities. "A few students can always be expected to fail a course." The summated rating for the KPG was 40 and the UNG 20. "Student abilities have greatly decreased in the last few years." The summated rating for KPG was 40 and the UNG 25.

A difference in how these two groups viewed training methods and design was also indicated. "Use of platform instruction is the best method of training." The summated rating for the KPG was 43 and the UNG 26.

Even though we have not applied extensive statistical procedures in our preliminary work it is indicated that instructors view concerns about ISD and performance based testing differently than training developers. The authors suggest the following strategies to minimize training turbulence that this difference may create in institutions implementing ISD:

1. Organizational Implementation Plan
2. Use of matrix organization
3. Using resistance to change as a resource to change
4. Formal end of course reviews

An organizational ISD implementation plan that has been approved and is fully supported by all senior staff members is a necessity. An implementation plan should include an extensive instructor orientation to the ISD process. Without such a plan implementation of ISD will likely be in a "muddling through" mode at best. In 1975, the Army's Infantry School at Ft. Benning developed an excellent ISD implementation plan consisting of 11 major activities carried out over a 15 month period. The result was an easier transition from traditional subject matter focused training to a performance based ISD system.

Since the introduction of ISD in the early 70's military schools have reorganized on the basis of function. This tends to further isolate the instructor from the developer. Matrix organization synonymous with project management in research and development would provide greater interaction between developers and subject matter experts. Organization on the basis of MOS with other required personnel would do a great deal toward reducing potential instructor resistance to the ISD process.

Labeling the instructor as a source of resistance to implementing ISD may create a serious psychological block in the minds of those responsible for implementing ISD programs. Instead of viewing the instructor as a source of resistance to the ISD process, it is recommended to cultivate the instructor group as a potential resource. There may be a very real and legitimate basis for what may seem to be voices of resistance. "ISD will take more time"; It won't work here"; We've always done it this way"; and other apparently negative positions may be sources of identifying better alternatives and improving original ISD planning.

Formal end of course reviews is an excellent means of dissemenating current information to instructors and course developers. Student abilities, student performance, what worked, what didn't work, are all items that should be included in an end of course review.

SUMMARY: This presentation has focused on the issue of instructor attitudes toward ISD and performance based testing. Our preliminary work indicates that a difference does exist between instructors and training developers and how each views certain concerns in the training environment.

The first phase of our project has resulted in the selection of statements for constructing a Likert Summated-Rating instrument. Formal item analysis will be necessary to validate the instrument and to apply it in an actual training environment. Anyone desiring to obtain a copy of this instrument may do so by a request to any of the authors.

Even though our preliminary work has been limited the issue of instructor attitudes toward ISD is a real one. Training organizations failing to recognize this potential problem will fine implementation of ISD in traditional subject-matter-based system difficult at best.

BIBLIOGRAPHY

(1) Knapper, Christopher, Kay: Evaluating Instructional Technology; Halstead Press; 1980.

(2) Steward, Judith and W.A. Love: "The Teacher's Response to Technology and the Individualization of Instruction", The Educational Technology Review Series; Educational Technology: 1977.

(3) Sink, Darryl L.: "Managing Personal Interation with Long Term Instructional Development Projects"; NSPI Journal; pp 13-14, Nov 1978.

(4) L.L. Thurstone: "Attitudes Can Be Measured"; American Journal of Sociology; pp 529-554, 1928.

(5) Liker, Rensis: "A Technique for the Measurement of Attitudes"; Archives of Psychology, 1932.

An Iterative Decision Method for Selecting Medical Tasks for Training

Kenn Finstuen, M.S., M.Ed., Ph.D.

Academy of Health Sciences, United States Army
Fort Sam Houston, Texas 78234

Abstract

This paper describes a group decision technique used by training policy managers of the Army Medical Department (AMEDD) to select and prioritize tasks for enlisted medical training programs. The method systematically quantifies a series of training decisions made by $k=5$ or 7 subject matter expert judges. During the first iteration judges make independent dichotomous decisions (1 = select versus 0 = nonselect) concerning $\underline{n}$ = 100 to 300 medical performance tasks for a specific military occupational specialty (MOS). Training decisions are analyzed via multiple linear regression procedures, and predicted task scores, goodness-of-fit, and inter-rater reliability measures are computed. Tasks are simultaneously rank ordered by task scores and degree of rater agreement and are displayed in a standard graphic format. The board is convened and the results from the first decision iteration (J1) are examined. Feedback results are used to direct group discussion. After discussion, judges render revised group decisions (J2) on disputed tasks. The judges then rate or rank the selected tasks in terms of combat criticality. Results are used in task analysis and medical training design and development.

"The views of the author are his own and do not purport to reflect the position of the Department of the Army or the Department of Defense."

Background

Decision making in a training environment can be viewed as a form of productivity. Within this approach, training decisions may be assessed along the dimensions of effectiveness and efficiency. To be effective, decisions should be accurate, be centered upon the appropriate issues, be understandable and be useful as an integrated product. To be efficient, decisions should be timely, be arrived at in an orderly, systematic manner, and be parsimonious in the expenditure of resources. The purpose of this paper is to describe the rationale for, and the dynamics associated with the Iterative Decision Method (IDM), a productivity based expert group decision-making technique developed by the Individual Training Division of the Academy of Health Sciences (AHS), Ft Sam Houston, TX. The method has been applied by an expert panel that consisted of an emergency room physician and nurse and 3 combat medics to select and prioritize tasks for the 91B30 Advanced Medical Specialist course (Carroll & Finstuen, Note 1). More recently the IDM was employed by an AHS Colonel's committee and an AMEDD General Officer board to prioritize medical combat deficiencies for mission area analyses (Finstuen, in press).

The IDM is a highly structured group judgment model designed to maximize the effectiveness and efficiency of decision making for an expert panel of 5 or 7 decision makers. The process consists of a nominal group phase in which members render independent task selection judgments (J1) about a well-defined list of potential medical training tasks. The J1 decisions are then statistically modeled using multiple linear regression equations. The J1 results are employed as feedback in the second face-to-face interaction group phase to arrive at a revised group judgment (J2). Each of the components of the IDM procedure have been carefully constructed to optimize decision-making effectiveness and to limit inefficient actions. The technology draws from several decision-making techniques and is based upon the research findings of over 70 small group interaction and productivity studies.

Productivity in Decision Making

Individual versus Group Decisions

Outputs from small groups have typically been found to exceed outputs of single individuals (Rosenberg, 1969). An extensive literature review of small

group and individual productivity from 1920 to 1957 (Lorge, Fox, Davitz & Brenner, 1958) presented evidence in favor of group versus individual productivity across a variety of performance tasks. For example, groups were shown to be more accurate than individuals in judgments of the weight of physical objects, social situations, and in the solution of complex problems. More recently Davis (1969) demonstrated that cognitive and intellectual task performance may be enhanced by group activity.

The need for collective decision making has long been recognized by the military as evidenced by numerous boards convened for personnel selection, promotion, disciplinary action, and budget and project planning. The goal of the IDM process is not to replace board actions, but rather to enhance the productivity of such actions.

Nominal versus Interactive Group Decisions

Although group decisions tend to be superior to individual decisions in terms of productivity, different approaches may be taken to arrive at a group decision. Are members of a board more effective and efficient in decision-making if they work individually or if they accomplish most of the work in face-to-face meetings? This question addresses the difference between nominal and interactive group structures. Comparative studies of decision making in small groups suggest that there are distinct advantages associated with the use of nominal versus interactive groups (Marquart, 1955). Differences in performance may be attributed to three main factors, viz., the judgmental difficulty associated with the objects of interest, characteristics of the group members who make the judgments, and the situation in which the judgments occur.

In reference to judgmental difficulty, both groups tend to function equally well for simple unitary task decisions (Kelley & Thibaut, 1969), but as decisions become more complex, interactive groups tend to perform more accurately and tend to be more satisfied with their performance (Faust, 1959; Hackman, 1968; Morris, 1966; Shiftlett, 1972). Many studies have investigated the moderating effects of task difficulty and interpersonal interaction upon performance and consequent satisfaction and have arrived at similar findings (Trow, 1957; Ewen, 1973; Bray, Kerr, & Atkin, 1978).

Several concerns separate the nominal from the interactive process in regard to the characteristics of group members. First, a nominal group maintains a higher degree of impartiality because members make their decisions individually. Independent action limits the amount of influence that board members may exert upon others (Van de Ven & Delbecq, 1971). Second, by discussion, members tend to stimulate thoughts that other members might not have if they work alone (Hall, Mouton, & Blake, 1963). Third, an interactive group benefits from a pooling of resources while the nominal does not. In terms of accuracy, face-to-face interaction provides opportunities for errors to correct themselves, for clarification of issues, and for an analysis of the logic behind member decisions (Delbecq, Van de Ven, & Gustafson, 1975). The pooling-of-abilities model has prompted a great deal of research (e.g., Goldman, 1965, 1966; Laughlin & Johnson, 1966; Shaw, 1971; Steiner, 1972). While the majority of these studies confirmed the obvious advantages of the pooling-of-abilities effect on decision making in interactive groups, findings also indicated that the effect was contingent on a high level of member ability (expertise). In groups composed of experts each member has unique specialized information, skills, and experiences that enhance collective decision making. In interaction groups composed of individuals with relatively low ability or knowledge of the issues; few, if any, gains were observed beyond the productivity of nominal group conditions.

In regard to situational effects on the productivity of decision making, the amount of time allowed for solutions appeared to be one of the major factors affecting both interactive and nominal group conditions (Restle, 1962).

Optimal group size. Another situational productivity factor involves the size of the interaction decision-making group. Steiner (1972) hypothesized that productivity generally increases with the size of the group up to a point where coordination and motivation decrements take over. In the case of coordination decrements, the larger the group, the greater will be the process loss due to the requirement that all members function in a concerted manner. For motivation effects, member effort declines as group size increases since the addition of more persons to the group decreases the individual amount of outcome rewards associated with the decision making. Research support for coordination effects has been consistent. Ziller (1957) found that decision accuracy increased 74% when the performance of one person was compared with a 3-person group. However, increments in productivity tended to be smaller as more people were added to the group, i.e., when the group was increased from 3 to 6 members, accuracy increased only 9 percentage points. Other researchers have reported that members experienced difficulty in the coordination of groups of more than 7 persons (James, 1951), and that as size increased above 7 that restraints against participation increased (Delbecq, et.al., 1975).

In addition to coordination effects, motivation is inversely affected by increases in the size of interactive groups. Slater (1958) has shown that in groups of from 2 to 7 members, that groups of size 5 were most satisfied with committee actions. Larger groups complained of inefficiency, while smaller groups became more concerned with interpersonal relations.

With respect to the optimal size for interaction decision groups, some investigators recommend a size of 5 (Bales, 1956; Slater, 1958), while others recommend a range from at least 5 to 7 (Delbecq, et.al., 1975; Hare, 1962; James, 1951). Groups of less than 5 probably lack the diversity of skills under the pooling-of-abilities model. Also, in groups of 5 or more it has been found that the opinions given are more carefully thought out before they are presented (Hare, 1962). These findings indicate that, for optimal productivity, interactive decision-making groups should consist of a least 5 but no more than 7 expert members. Further, the use of an odd number is recommended to circumvent the possibility of a deadlock.

In summary, the evidence from the research literature indicates that 1) collective decisions are more productive than decisions made by a single individual, 2) that nominal groups are most useful for making unitary task decisions and for maintaining impartiality, and 3) that interactive groups of 5 or 7 experts tend to be more satisfied and more productive in making complex decisions.

Maximizing Decision Productivity For Medical Expert Boards

Many special purpose techniques exist for modeling expert judgments and decision making such as the Delphi survey technique (Turoff, 1970), and the nominal group technique--NGT (Delbecq, et.al., 1975; Vroman, 1975). Judgment models can be differentiated by the type of group structure used (nominal vs interactive), and the types of judgments employed for making decisions among items or tasks. Nominal group results can be used two ways. First, decision information may be used directly as an end product. Second, results can be input to an interactive group decision. Results from Delphi and NGT research have shown that group decisions benefit from the sequence of nominal and interactive group actions. In this regard nominal group judgments may be

viewed as a form of front-end-analysis--FEA (Harless, 1975) for the interactive round of decision making.

The following task selection model was developed to maximize the judgment productivity of AMEDD expert boards (see Figure 1). Each of the components of the procedure have been carefully structured to optimize decision-making effectiveness and to limit inefficient actions. Analogcus to the medical model, independent judgments from nominal groups, defined as J1, may be viewed

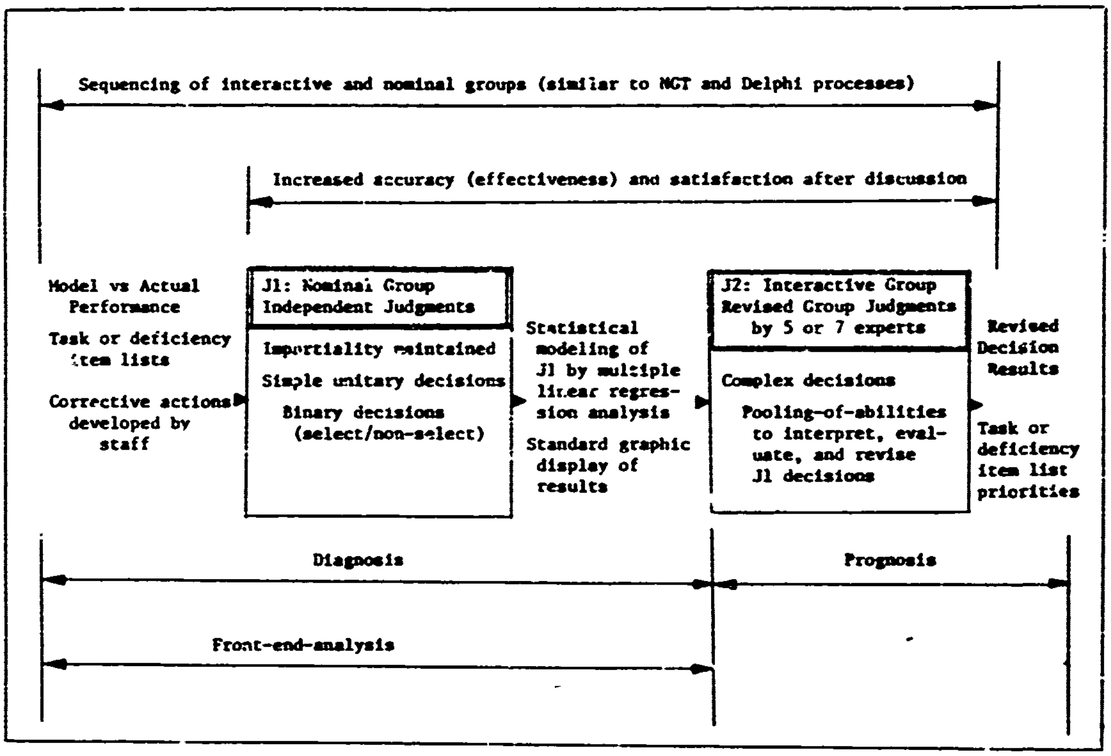


Figure 1. Diagram of the Iterative Decision Method for maximizing the productivity of expert board judgment and decision-making actions. The model is an integration of several decision-making techniques, small group productivity concepts and research, and applied statistical analysis.

as an aid to diagnosis, that is they provide information about the present state of a given decision-making situation. Revised group judgments defined as J2, involve a prognosis for corrective actions to be taken. As shown, nominal (J1) and interactive (J2) groups are alternated. In addition to the separate advantages of using both types of groups, the combined use of a nominal group first and a discussion group second results in an increase in effectiveness. The J1 results are used as feedback during the J2 phase of decision making, and tend to increase the accuracy of decisions and to decrease the range of judgments after discussion (Delbecq, et.al., 1975; Thorndike, 1938).

To increase the efficiency of the process, task lists are developed by a FEA staff before the decision-making process begins. Unlike the NGT and Delphi, the FEA frees the expert panel from the job of identifying judgment items (i.e. potential training tasks) prior to making decisions of training requirements.

The IDM Process For Medical Training Decisions

During the J1 round of judgments, 5 or 7 medical expert decision-makers impartially provide simple decisions in reference to the judgment items. For

the selection of tasks for training, decisions are either yes (coded 1) or no (coded 0). Lists consist of 100-300 medical tasks grouped into 15-25 duties per MOS. While simple averages help to differentiate among tasks, explicit measures of expert agreement, computed from multiple linear regression equations, provide a comprehensive and efficient picture of the J1 decision results. Predictive information consists of both task and judge variables that are statistically analyzed with a series of special application APL computer programs. In addition to standard descriptive statistics, i.e. task averages and standard deviations, three statistical indices are computed for each duty list. For each duty, indices reflect the goodness-of-fit for a group equation that expresses individual judgments as a function of a set of binary task and expert predictor variables, an index of the inter-rater reliability, and an _F_ test which expresses the results of testing the hypothesis of task mean differences. In addition, standardized graphic displays of duty results allow the experts to efficiently direct the group discussion and to focus on disagreements which merit attention. The only reason that agreed upon tasks are discussed is to identify the rationale used to arrive at selection decisions. While this information might be interesting, it is secured at the expense of time which would be better directed to the pressing problem areas. For the sake of efficiency, expert positions are identified only when they are associated with disagreements. With this form of decision making there are no correct or incorrect opinions, however, the probability of 100% consensus for all J1 decisions is remote. The objective of the process is to have the group arrive at an acceptable level of agreement; it is not necessary that 100% consensus be obtained at J2.

Once the revised J2 selection decisions have been made, the experts either rank or rate the tasks in terms of combat criticality. The result is a prioritized list of medical tasks to be developed for specific MOS training.

Reference Note

1. Carroll, T., & Finstuen, K. US Army Advanced Medic (91B30) Training: An iterative decision method application. Paper to be presented at the 24th Annual Conference of the Military Testing Association, San Antonio, Nov 1982.

References

Bales, R. In conference. Harvard Business Review, 1954, 32, 44-50.

Bray, R., Kerr, N., & Atkins, R. Effects of group size, problem difficulty, and sex on group performance and member reactions. Journal of Personality and Social Psychology, 1978, 36, 1229-1240.

Davis, J. Group performance. Reading, Mass.: Addison-Wesley, 1969.

Delbecq, A., Van de Ven, A., & Gustafson, D. Group techniques for program planning: A guide to nominal group and Delphi processes. Glenview, Ill.: Scott, Foresman, and Company, 1975.

Ewen, R. Pressure for production, task difficulty, and the correlation between job satisfaction and job performance. Journal of Applied Psychology, 1973, 58, 378-380.

Faust, W. Group versus individual problem-solving. Journal of Abnormal and Social Psychology, 1959, 59, 68-72.

Finstuen, K. Prioritization of medical combat deficiencies: Application of the iterative decision method (AHS Monograph 82-1). Ft Sam Houston, Tx.: Academy of Health Sciences, Directorate of Training Development, Sep 1982.

Goldman, M. A comparison of individual and group performance for varying combinations of initial ability. Journal of Personality and Social Psychology, 1965, 1, 210-216.

Goldman, M. A comparison of group and individual performance where subjects have varying tendencies to solve problems. Journal of Personality and Social Psychology, 1966, 3, 604-607.

Hackman, J. Effects of task characteristics on group products. Journal of Experimental Social Psychology, 1968, 4, 162-187.

Hall, E., Mouton, J., & Blake, R. Group problem-solving effectiveness under conditions of pooling versus interaction. Journal of Social Psychology, 1963, 59, 147-157.

Hare, A. Handbook of small group research. New York: The Free Press, 1962.

Harless, J. An ounce of analysis. Newnan, Ga.: Performance Guild, 1975.

James, J. A preliminary study of the size determinant in small group interaction. American Sociological Review, 1951, 16, 474-477.

Kelley, H., & Thibaut, J. Group problem solving. In G. Lindsey and E. Aronson (Eds). The handbook of social psychology (2nd ed) Vol. 4. Reading, Mass.: Addison-Wesley, 1969. 1-101.

Laughlin, P., & Johnson, H. Group and individual performance on a complementary task as a function of initial ability level. Journal of Experimental Social Psychology, 1966, 2, 407-414.

Lorge, J. Fox, D., Davitz, J., & Brenner, M. A survey of studies contrasting the quality of group performance and individual performance from 1920 to 1957. Psychological Bulletin, 1958, 55, 337-372.

Marquart, D. Group problem solving. Journal of Social Psychology, 1955, 41, 103-113.

Morris, C. Task effects on group interaction. Journal of Personality and Social Psychology, 1966, 4, 545-554.

Restle, F. Speed and accuracy of cognitive achievement in small groups. In V. Criswell, H. Solomon, & P. Suppes (Eds). Mathematical methods in small group processes. Stanford: Stanford University Press, 1962, 250-262.

Rosenberg, S. Mathematical models of social behavior. In G. Lindsey and E. Aronson (Eds). The handbook of social psychology (2nd Ed), Vol 1., Reading Mass.: Addison-Wesley, 1969, 179-244.

Shaw, M. Group dynamics: The psychology of small group behavior. New York: McGraw-Hill, 1971.

Shiftlett, S. Group performance as a function of task difficulty and organizational interdependence. Organizational Behavior and Human Performance, 1972, 7, 129-139.

Slater, P. Contrasting correlates of group size. Sociometry, 1958, 21, 129-139.

Steiner, I. Group process and productivity. New York: Academic Press, 1972.

Thorndike, R. The effect of discussion upon the correctness of group decisions when the factor of majority influence is allowed for. Journal of Social Psychology, 1938, 9, 343-362.

Trow, D. Autonomy and job satisfaction in task-oriented groups. Journal of Abnormal and Social Psychology, 1957, 54, 204-209.

Turoff, M. The design of a policy Delphi. Technological Forecasting and Social Change, 1970, 2.

Van de Ven, A., & Delbecq, A. Nominal versus interacting group processes for committee decision-making effectiveness. Academy of Management Journal, 1971, 14, 203-212.

Vroman, H. An application of the nominal group technique in educational systems analysis. Educational Technology, 1975, 51-53.

Ziller, R. Group size: A determinant of the quality and stability of group decisions. Sociometry, 1957, 20, 165-173.

AD P000869

SELECTION OF INTELLIGENCE ANALYSTS

M. A. Fischl

and

Arthur C. F. Gilbert

U.S. Army Research Institute for the
Behavioral and Social Sciences

The Defense Intelligence Agency (DIA) is responsible for keeping the Joint Chiefs of Staff informed of activities of potential military consequence world-wide. A very important element in performance of this mission is the uniformed and civilian staff of individuals who analyze social, political, economic, and strategic information about assigned parts of the world.

Civilian intelligence analysts are selected for their jobs largely on the basis of academic record and prior work history. Virtually all are college graduates, many with graduate training and degrees, and many are former uniformed military personnel. This paper describes a recent investigation into the feasibility of improving the process of selection of civilian intelligence analysts, through adding the use of tests of the aptitudes and skills required in the job.

Method

The method employed followed a standard test-development paradigm. Job analysis identified personal characteristics important to analyst success, an experimental battery of tests to measure the characteristics was selected, and it was administered to a sample of recently-hired incumbents for whom job performance information was also obtained. Multiple regression analyses weighted the tests, which were then cross-validated on holdout portions of the sample.

Job Analysis

Discussions were held with members of the DIA staff, to learn the nature of the job performed by intelligence analysts and apparent causes of success and failure on the job. Additional information about the analyst job and characteristics judged important for its successful performance was obtained from personal interviews with 14 incumbent intelligence analysts and from critical incident questionnaires on which 20 supervisors of analysts provided descriptions of positive and negative critical incidents, with explanation of the personal qualities responsible for each incident.

Experimental Test Battery

On the basis of the job analysis, a picture emerged of personal attributes important in the intelligence analyst job. Table 1 presents these attributes and the commercial tests selected to measure them and serve as potential predictors of analyst success.

Table 1
Experimental Predictor Variables
and Their Tests

Variable	Test
High Level reasoning ability	Watson-Glaser Critical Thinking Appraisal 1/
Inductive reasoning	Comprehensive Ability Battery: Subtest 6, Inductive Reasoning 2/
Intellectual flexibility	Comprehensive Ability Battery: Subtest 15, Spontaneous Flexibility
Writing skill	Flanagan Industrial Tests: Subtest 6, Expression 3/
Memory	Flanagan Industrial Tests: Subtest 12, Memory
Intellectual curiosity	Gordon Personal Inventory: Original Thinking Scale 1/
Deliberateness, carefulness	Gordon Personal Inventory: Cautiousness Scale
Interpersonal skill	Gordon Personal Inventory: Personal Relations Scale
Achievement motivation	Gordon Survey of Personal Values: Achievement Scale 3/
Self-discipline	Gordon Survey of Personal Values: Orderliness Scale
Perserverance	Gordon Personal Profile: Responsiblity Scale 1/

1/ New York: The Psychological Corporation

2/ Champaign, IL: Institute fo_ Personality and Ability Testing

3/ Chicago: Science Research Associates

Subjects and Procedure

The experimental battery was administered in a 3-hour session to 64 intelligence analysts who had been employed at DIA for periods ranging from 1 to 24 months. The mean experience level was just under 12 months, and the sample was approximately 2/3 male, 1/3 female. All but 3 members of the sample were caucasian. These 64 analysts were the most recently hired by DIA.

Immediately after the testing session a DIA staff member met individually with the supervisors of the 64 analysts to administer performance rating forms. At that time the supervisors were informed that the ratings were ad hoc and for research purposes only, would not appear on a personnel record, and, when completed, were to be transmitted in sealed envelopes directly to the research organization outside of DIA. Candor and accuracy were encouraged, it being pointed out that there was no risk to any employee but potential great benefit to the Agency. A copy of the rating form appears as Figure 1.

ANALYST RATING FORM

Date:________________

Analyst's Name (Print) ____________________
Last First MI

Rater's Name (Print) ____________________
Last First MI

Length of time you have known this analyst ____________ (Months)

INSTRUCTIONS

COMPARED TO ALL ANALYSTS YOU HAVE KNOWN, using the rating scale below, blacken one box to indicate your appraisal of this analyst's performance. In making your rating of performance, consider the analyst's demonstrated performance relative to the performance of all other analysts you have known at his/her stage of experience.

Bottom 10% (Marginal)	Next higher 20%	Middle 40% (Average)	Next higher 20%	Top 10% (Outstanding)
1	2	3	4	5

Figure 1. Performance appraisal instrument

The full sample of 64 intelligence analysts was randomly divided into two half samples, and independent stepwise multiple regression analyses were performed on each. The regression weights emerging from analysis of half-sample _A_ were utilized to compute a score for each member of half-sample _B_, and that score distribution was correlated with the distribution of criterion ratings for these individuals. Similarly, the regression weights emerging from analysis of half-sample _B_ were utilized to compute a score for each member of half-sample _A_, and that score distribution was correlated with the distribution of their criterion ratings.

The validation procedure will be recognized as standard double cross-validation, yielding two regression equations and two validity coefficients. From that point, judgment was utilized to integrate the two solutions— that is, to select the final test battery-- and to arrive at a single best estimate of the criterion-related validity of that battery.

Results

The most valid test was the test of Expression, correlating 0.55 with the criterion in half-sample _A_ and 0.37 in half-sample _B_. Of equal validity to Expression in half-sample _B_ was the Critical Thinking Appraisal, and this test also correlated 0.36 with the criterion in half-sample _A_.

Addition of tests resulted in five-variable solutions in both half-samples. In half-sample _A_ this solution was:

$$Y = 2.144 + 0.054 \text{ Expression} - 0.036 \text{ Orderliness} + 0.015 \text{ Memory} + 0.015 \text{ Spontaneous Flexibility} + 0.010 \text{ Critical Thinking}$$

In half-sample _B_ the five-variable solution was:

$$Y = 1.36 + 0.046 \text{ Expression} + 0.026 \text{ Critical Thinking} - 0.025 \text{ Orderliness} + 0.022 \text{ Memory} + 0.013 \text{ Spontaneous Flexibility}$$

When these five-variable regression equations were cross-validated in the opposite half-samples, the resulting correlation coefficients were 0.60 and 0.38.

Note was taken that Orderliness had a negative regression weight in both solutions. For operational application the use of negative weights was judged highly undesirable, and the Orderliness scale was deleted from further consideration. At the same time, more careful examination of the Spontaneous Flexibility test disclosed that it could not be scored by a non-professional, that careful subjective judgment was needed. This second administrative concern disqualified the test of Spontaneous Flexibility. The remaining three variables were common to both regression equations, and the only tasks remaining were to derive a new set of weights for a three-variable equation, and to estimate the validity of the three-test battery. These tasks were performed in each half-sample, crossed on the other, and also in the full sample of 64 cases. Table 2 presents the outcomes.

Table 2
Regression Weights and Validity Coefficients
for the Three-Test Battery

Weights	Validity Coefficient
0.069 Expression + 0.011 Critical Thinking + 0.009 Memory	0.42, N_A = 32
0.045 Expression + 0.032 Critical Thinking + 0.003 Memory	0.54, N_B =32
0.057 Expression + 0.021 Critical Thinking + 0.007 Memory	0.50, N =64

Discussion

The three-test battery consisting of the Watson-Glaser Critical Thinking Appraisal, and the Memory and Expression subtests from the series of Flanagan Industrial Tests, requires about 1½-hours for administration and is scorable by a clerk using stencil overlays.

If simple unit weights--i.e., the sum of raw scores on the three tests--are employed in operational use of the battery, the counterparts to the validity coefficients shown in Table 2 become respectively 0.41, 0.48, and 0.44, all significant at $p \leq 0.01$.

Analysis to detect any adverse gender impact disclosed no difference in test battery scores of the women and the men in the sample. Using the simpler weights, for which the total possible score is 150, the women's mean was 105 and the men's mean was 104.

An alternative to the unit weighting procedure might be differenti. whole-number weighting of the tests of the battery. Inasmuch as the regression analyses weighted Expression between 1½ and 7 times as heavily as Critical Thinking and between 7 and 15 times as heavily as Memory, a set of weights in the ratio 5:2:1 or 6:2:1 might be superior. Its use was not investigated.

On the basis of the investigation performed, it appears that a relatively short battery of easily administered and scored tests can appreciably improve the procedure for selection of intelligence analysts.

Acknowledgment

The authors acknowledge the assistance of Ms. Donna Angle in this activity.

AD P000870

DEVELOPMENT OF TANK PLATOON GUNNERY TACTICAL STANDARDS THROUGH SIMULATION[1]

by
Patrick Ford
Roy C. Campbell
(HumRRO)

start

Highly realistic training exercises are not necessarily suitable testing means. The difficulties are highlighted in the Army Tank Gunnery Table IX - Platoon Battlerun, which is a live fire, free play exercise. As outlined in FM 17-12-2, Tank Gunnery, M60 Ch. 2, Table IX is scored objectively by an assessment of target hits, engagement times and ammunition conservation. More subjective evaluations are made of the platoon's tactics which involve use of terrain, cover and concealment, and movement techniques. The purpose of the work described in this paper was to identify meaningful tactical measures that would discriminate between qualified and unqualified tank platoons and improve the objectivity of such measures.

The immediate procedural problem to overcome was that, like most tactical exercises, Table IX is expensive and time consuming to conduct even under dry fire or subcaliber fire conditions. Table IX is especially complex because of the impact of variables such as weather, vehicle breakdowns, safety restrictions, minor terrain variations and individual tank commander capabilities. While such variables must be dealt with during any tactical exercise, considering them at this phase would have obscured focus on identifying and expressing candidate tactical measures.

To overcome the problems of cost and complexity, the general approach was to have platoon leaders (PL) conduct a simulated offensive mission on either a map or terrain boards. The first study was conducted on a pictomap of Fort Carson, Colorado marked with grid lines and contour lines. The second study was conducted on two terrain boards. One covered the battlerun area at Fort Carson, Colorado; the other covered the battlerun area at Fort Knox, Kentucky. All three were in the approximate scale of 1:2500.

The number of distinct threat arrays was reduced from the FM 17-12-2 Table IX requirement of eight to four. That is still more than a platoon would challenge without additional support. The enemy locations were selected at the start of data collection and were the same for each subject.

With both formats a PL was given an operations order for movement to and occupation of a terrain objective. He then planned his movement including preplotted artillery targets. After reporting his preplots the PL moved scale model tanks along the route he had selected. When the platoon came within the effective range of an enemy position, a controller announced that the platoon had been fired upon, who had received the fire, and where the enemy was located. The PL options for reacting to the fire included maneuvering,

---

[1]The work described here was performed under contract with the U.S. Army Research Institute for the Behavioral and Social Sciences. G. Gary Boycan was the Technical Representative. George Wheaton, American Institutes for Research, directed the project. The opinions expressed are those of the authors.

firing, and calling for indirect fire. The delay for delivering indirect fire varied by the proximity of the preplots. All other requests for support including reinforcement were denied.

The movement and engagement rules were based on protocols developed by Medlin[1] for combined arms engagements. Each move represented about one minute of real time. Distance was controlled by terrain; maximum distance per move varied from 50 meters through wooded areas to 250 meters over open terrain. Casualties were not assessed on the platoon.

Study One: Experienced Vs. Inexperienced

Subjects

Two studies were conducted. The first study was conducted on the pictomap and compared the performance of experienced PL with inexperienced PL. The experienced group included three 1LT (O-2) and one SFC (E-7). All had participated in battleruns as Platoon Leader or Platoon Sergeant. The inexperienced group contained five students (O-1) in the Armor Officer Basic Course. All had completed classwork, terrain board practical exercises, and REALTRAIN field exercises on offensive operations. Three had fired Table IX, one as a Platoon Leader.

Results

The results identified two promising aspects for evaluating the PL and the platoon: indirect fire planning and movement techniques.

Table 1 shows the distribution of four categories of effective preplots for indirect fire. There appears to be a range of ability with indirect fire. Almost everyone planned for fire on the objective, which is a reasonable minimum requirement. The experienced group had a clear edge in ability to identify potential enemy locations.

Table 1
Indirect Fire Preplots

Subject		Preplots			
		On Objective	In Front Of Objective	On Enemy*	Within 400 M Of Enemy*
Exper.	1	1	1	1	1
	2	1	1	1	1
	3	1			1
	4	1			2
Inexper.	1	1			
	2	1			
	3	1			1
	4	1			
	5		1	1	1

*Preplots were made prior to problem play without knowledge of actual threat locations. This table shows the number of preplots that matched threat locations. The 400 m category requires the minimum adjustment delay.

[1]Medlin, Steven M. Behavioral Forecasting for Realtrain Combined Arms (Technical Report 365). U.S. Army Research Institute for the Behavioral and Social Sciences, 1979.

Table 2
Number of Times Prior To Assault
That Both Sections Moved During the Same Turn

Group	Range of Total Moves	Mean	Range of Simultaneous Moves	Mean
Experienced	10-27	22	0-3	2.25
Inexperienced	20-29	25.6	4-5	4.4

In terms of movement techniques, all participants traveled basically in bounding overwatch; one tank section[1] moved while the other occupied an overwatch position where they could engage likely enemy. This is the doctrinally correct technique for this type of mission. Still, there were differences in the way the groups implemented the technique. One difference was that the inexperienced group were less likely to use bounding overwatch consistently that the experienced group. Table 2 shows the number of moves where there was no effort at overwatch. The inexperienced players averaged about twice as many nonoverwatch (simultaneous) moves. The second difference in movement techniques related to the size of the bounds (that is, the distance the maneuvering section travels before the overwatch section displaces). As shown in Table 3, the average bound for experienced subjects was about 45% longer than for the inexperienced subjects. The apparent difference between the groups is promising, especially for continuous measures of movement techniques (rather than GO/NO GO measures).

Table 3
Distance of Bounds

Group	Range of # of Bounds	Mean	Range of Mean Length (Meters)	Mean
Experienced	2-4	2.75	425-850	591
Inexperienced	2-7	4.2	350-625	414

Study Two: Expert Cover and Concealment

The second phase of the data collection was conducted on terrain boards. The primary focus of this part of the data collection was to explore the suitability of numerical values of cover and concealment.

Subjects

Three groups of experts on platoon tactics worked through an offensive mission on the Fort Knox terrain board and on the Fort Carson terrain board. Groups were from Armor School staff: Directorate of Training Developments (DTD), Command, Staff and Doctrine Department (C&S), and Weapons Department (WPN). There were three experts in each group. The experts agreed among

[1]The organization given the participants was the five tank platoon equipped with M60A1 tanks with Add On Stabilization. Normal employment dictates controlling the platoon by section. The Heavy Section normally consists of the PL tank and two other tanks; the Light Section of the platoon sergeant's tank and one other tank. With this division a variety of overwatch and maneuver combinations are possible depending on the terrain, mission and tactical situation as outlined in FM 71-1, Armor Operations.

themselves on the general concept of the operation. For each move, one of the experts proposed a move and the others responded until they reached a consensus. Otherwise the same movement and engagement rules applied as for the pictomap.

Procedure

After the missions were completed, the routes were plotted again on the boards. The four enemy locations in each exercise were the locations for enemy weapons. The segments where a tank section was in an exposed area within effective range of the enemy weapon were measured. A section was considered to be exposed if there was line of sight from the enemy location to the section. Movement through wooded areas or along woodlines was considered to be concealed. If the PL coordinated the delivery of smoke to mask a move, exposure was calculated both without smoke and with maximum effectiveness for the smoke. The results reported here are for exposure without smoke.

Results

The central problem was to establish a standard for exposure that could be applied to several battlerun locations. We had hoped to set a combat referenced standard where any tank that was exposed to an observer at an enemy position long enough for effective fire would be considered dead. Pilot studies showed this approach has limited benefits for Table IX for two reasons. First, since Table IX is live fire, observers could not be located at the enemy positions. Second, at least using the size threat in FM 17-12-2 Table IX, all tanks would have been exposed long enough to be killed at least once. This approach would be comparable to a multiple-choice question where all alternatives are incorrect.

In the absence of a suitable combat referenced standard, we examined three candidate indexes. Meters exposed to the enemy is the basis for all indexes.

Percent of Movement Exposed. The first index is percent of movement exposed. The obtained exposure is divided by total distance traveled. If a section traveled 4000 meters and was exposed for 2000 meters, its index would be 50%. The percent of movement exposed indexes are shown in Table 4.

With one exception, the indexes are very similar, an encouraging result since each group was considered expert. The exception resulted from a limitation in the terrain board medium where WPN moved along the edge of the board in areas that were not exposed to the enemy positions on the board but would have been exposed to likely enemy positions on adjacent terrain.

Table 4
Percent of Movement Exposed

Group	Fort Knox	Fort Carson
DTD	59%	78%
C&S	63%	79%
WPN	63%	54%

This index had been expected to vary depending upon the terrain. Since a battlerun conducted on flat, sparsely wooded terrain--such as Fort Carson--offers fewer opportunities for cover and concealment than a battlerun conducted on rolling, moderately wooded terrain--such as Fort Knox--platoons were expected to be exposed for a higher proportion of their movement on the Fort Carson battlerun than on the Fort Knox battlerun. And, as shown in Table 4, there was a tendency toward higher exposure for Fort Carson than for Fort Knox. Because of this tendency, an "expert" level would have to be established for each location. A generalizable minimum standard is not apparent.

Percent of Baseline Exposure. This figure is derived from a plot of the most direct trafficable route from the assembly area to the objective. Exposure on this route indicates the amount of cover and concealment that would be obtained by accident. For example, the Fort Knox battlerun baseline exposure is 3200; a section that was exposed for 2000 meters would obtain an index of 63.

Table 5
Percent of Baseline Exposure

Group	Fort Knox	Fort Carson
DTD	90%	85%
C&S	78%	86%
WPN	90%	58%

The results of this analysis are shown in Table 5. Two characteristics should be noted:

- All groups chose routes that exposed them less than the baseline route. This suggests a logically appealing minimum standard: A platoon should select a route that provides enough cover and concealment to match or reduce exposure along the most direct route.

- The indexes do not get higher as the amount of cover and concealment available decreases. In fact, two groups obtained a lower index on the open Fort Carson terrain than on the rolling, moderately vegetated Fort Knox terrain. This suggests two possibilities. First, the concern for increasing cover and concealment may intensify as the amount of available cover and concealment decreases. Or, the shortage of effective enemy positions makes it easier for the platoon to protect itself from the actual positions.

Percent of Straightline Exposure. This index is based on plots of straightline routes between each position for a section. The amount of exposure for these bounds is summed and used as the denominator. If a section's actual route exposes the tanks for 2000 meters and the straightline route would have exposed them for 2400 meters, the section's index is 83.

The rationale for the index of straightline exposure is similar to the rationale for the index of baseline exposure. The intent is to find the gain in cover and concealment over what would be obtained by accident. The difference is that the hypothetical route for the straightline index is based on movement between overwatch positions on the assumption that a Platoon Leader might be willing to trade increased exposure for good overwatch positions. The question was how much will the platoon maneuver to decrease exposure when moving between the positions?

Table 6
Percent of Straightline Exposure

Group	Fort Knox	Fort Carson
DTD	149%	100%
C&S	96%	95%
WPN	96%	98%

As Table 6 indicates, the answer appears to be "not much." With one exception the exposure of the route is very close to the straightline exposure. The exception is that the exposure of the DTD route on Fort Knox is considerably higher than the straightline exposure. The main reason for the increase is that one section avoided wooded areas even though the areas would have decreased exposure, with the justification that wooded areas would also have decreased speed of movement and provided inadequate fields of fire during the bound.

Discussion

Ability to take advantage of cover and concealment is an important aspect of a platoon's overall tactical ability. The route a platoon travels is an important factor in the amount of cover and concealment that is available. But giving a score based on the route selected presents enough problems that each group responsible for a battlerun should analyze the particular situation before committing to any score.

The major problem to overcome in an evaluation is that it is very difficult to say that any route from the assembly area to the objective is unacceptable because it exposes the platoon too much. The routes on the terrain board exercises indicate that selecting a route involves tradeoffs among cover and concealment, fields of fire, and trafficability. The weight a PL assigns to any factor depends on elements such as knowledge of the enemy situation, the mission, the time available and the personality of the leader. The conditions for a live-fire battlerun reduce concern for cover and concealment even more. The understandable desire to focus on firing at and reacting to targets tends to make acquisition of fields of fire the dominant concern. There is also often pressure to evaluate several platoons during a day. Administrators may encourage, or even require, platoons to take more exposed but faster routes than normal. Because so many factors vary, trying to evaluate a platoon in terms of cover and concealment for the complete route is often comparable to asking a multiple choice question where all alternatives are correct.

There may be segments of a route that all platoons must cross, and that segment may provide attractive, incorrect alternate routes. If so, evaluating cover and concealment during that segment would be meaningful. Such areas are most likely between the line of departure and the first target array.

Recommendations for Evaluating Cover and Concealment. The analyses of routes on terrain board exercises suggest two recommendations that may benefit a unit that wants to evaluate cover and concealment in a Table IX battlerun.

1. Determine whether the mission and the terrain support evaluating a platoon's level of qualification in terms of cover and concealment. This decision involves two questions:

   - Is there an occasion during the mission when a qualified PL or Section Leader would make cover and concealment his primary concern?
   - When that occasion occurs, does the terrain offer correct routes and incorrect routes that the platoon or section will be allowed to take? If tanks are not allowed in the "incorrect" areas (because of safety fan or trafficability concerns), there would be no difference among platoons regardless of their level of qualification.

   Both questions should be answered "yes" before evaluators are committed to evaluating cover and concealment of a route segment.

2. If a cover and concealment index is used, establish chance and expert exposure. The standard for a minimally qualified crew or platoon is somewhere between amount of cover and concealment a platoon would obtain by chance and the amount an expert platoon would obtain. Of the indexes presented here, baseline exposure appears most valuable.

Conclusion

Because of problems with cost and troop support the benefits of the dimensions identified and the approaches to express the dimensions have not been validated. We are convinced though that the simulation technique used is promising for determining standards for tactical exercises. The approach elicits the type of decisions required for combat while presenting four advantages over deriving standards from "dry" battleruns:

- Cost, in terms of personnel, equipment, and time, is much lower.
- Researchers can isolate specific tactical conditions for study and replication.
- External variables do not control tactical decisions.
- Researchers can obtain more accurate transcripts of movement. This allows a trial and error approach to identifying meaningful measures.

Predicting Subsequent Service Potential for Former Army Prisoners

Dr. Lawrence J. Fox
Research Psychologist
Research & Evaluation Division
U.S. Army Retraining Brigade
Fort Riley, Kansas

CPT James M. Georgoulakis
Chief
Research & Evaluation Division
U.S. Army Retraining Brigade
Fort Riley, Kansas

Recent studies completed at the U.S. Army Retraining Brigade, a seven-week program for Army prisoners, at Fort Riley, Kansas, have provided comprehensive personality profiles of the Army's prisoner population (Georgoulakis & Fox, 1982). Administering the Sixteen Personality Factor Questionnaire (Cattell, Eber, & Tatsuoka, 1970) to 550 prisoners entering the program, Fox (1980) identified 10 scales with significant differences between those individuals who later graduated and their counterparts who failed to complete the program. Georgoulakis (1982), with a battery of 7 scales from the California Psychological Inventory (Gough, 1957), 2 scales from the Edwards Personal Preference Schedule (Edwards, 1959), Rosenberg's (1965) Self-Esteem Scale, and Hudson's (1974) Index of Self-Esteem, found significant differences between graduates and non-graduates on 6 scales.

The results of the two studies are consistent, and suggest that the graduates of the retraining program have more self-control, a better sense of personal responsibility, and are more sociable than those who fail to complete the program. Non-graduates, on the other hand, tend to be more independent, aggressive, and more careless or indifferent. It is important to note that these differences exist a priori, and are not causal effects of the program. This suggests that individuals who complete the training successfully may well have personalities better suited to the specific requirements of the Retraining Brigade program, and probably to the Army environment in general, than their non-graduate counterparts.

Until only recently, individuals selected for graduation (and further military service) were identified solely by a consensus of opinion on the part of their training team cadre. The purpose of the present study is to determine the extent to which personality measures, employed as independent variables, can predict graduation from the Retraining Brigade and the quality of performance during subsequent assignments. A parallel purpose of the study is to determine whether military and personal history data, available from conventional military records, offer a pool of potentially superior predictor variables.

Methodology

Since the two studies were conducted with different samples, two series of analyses were required. In each case, the various personality dimensions were entered as the independent (predictor) variables into a discriminant function analysis in order to predict graduation (versus an administrative discharge) at the Retraining Brigade. Next, 10 military/personal history variables, collected from the same samples, were employed in precisely the same manner, and the results compared.

Of the 550 prisoners to whom the 16PF was administered, 263 graduated and were returned to subsequent duty assignments with new units. After a three-year folow-up, Separation Program Designators, collected from DD Form 214, were recorded for each of the graduates. Success was defined as an Honorable Discharge upon completing military service, while failure in the subsequent assignment was defined as a General Discharge, a discharge under other than Honorable conditions, additional military or civilian confinement, and those individuals dropped from rolls (DFR). Using these two categories as the dependent variable, the 16PF standard scores and the 10 background variables were each entered into discriminant function analyses in order to determine the extent to which subsequent duty performance could be predicted from data collected upon entering the program.

In all cases, variables were entered into the discriminant functions concurrently (rather than stepwise) in order to enhance direct comparisons. A total of 6 discrimant functions were computed, utilizing computer programs from the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975).

Findings

a. Predicting Graduation or Discharge at the Retraining Brigade.

The discriminant functions in Table 1 represent linear combinations of the predictor variables which best distinguished between graduates (subsequently returned to new units) and those who were discharged after failing to complete the Retraining Brigade program successfully. The coefficients (interpreted in the same manner as factor weights) indicate the extent to which each variable contributed to differentiation between the two groups.

Both personality instruments produced discriminant functions which appear logically consistent. The 16PF described graduates as controlled (Q3), emotionally stable (C) and persevering (G), while portraying those who failed to graduate as aggressive (E) and independent (Q2). The CPI scales indicated that graduates tended to have a greater degree of self-acceptance (Sa), were more sociable (So), more responsible (Re) and had more self-control (Sc) than those who failed to complete training. Non-graduates had a greater need for autonomy, on the EPPS, and less self-esteem, on the Rosenberg (1965) scale.

Table 1

Predicting Graduation/Discharge from Training

16 Personality Factors (N=550)	Background Variables (N=550)
-.338 Q3 (Controlled)	-.600 Offense Category
-.313 C (Stable Emotionally)	-.560 Highest Pay Grade
-.312 G (Persevering)	-.352 Number of Dependents
.259 E (Assertive, Aggressive)	.342 Marital Status
.250 Q2 (Independent)	-.244 Court-Martial Category
-.243 H (Socially Bold)	.194 Race
.228 F (Happy-Go-Lucky)	.182 Months' Remaining Service
-.199 A (Outgoing, Friendly)	-.169 Age
-.182 Q4 (Tense, Frustrated)	

Group Centroids

-.394	Graduates	-.332
.361	Non-Graduates	.304

11 Selected Scales (N=100)	Background Variables (N=100)
.743 Self-Acceptance (CPI)	.588 Education Completed
.410 Socialization (CPI)	-.445 Highest Pay Grade
-.388 Social Presence (CPI)	-.443 Marital Status
.388 Self-Esteem (Rosenberg)	.384 Court-Martial Category
-.379 Need for Autonomy (EPPS)	.261 Number of Dependents
-.264 Dominance (CPI)	-.246 Offense Category
.178 Self-Control (CPI)	-.200 Age
-.114 Index of Self-Esteem (Hudson)	

Group Centroids

.281	Graduates	.344
-.281	Non-Graduates	-.352

Predictors	Eigenvalue	Wilks' Lambda	Graduates Correctly Predicted	Failures Correctly Predicted	Predict. Validity
16 PF (N=550)	.143	.874	67.3%	67.9%	.676
Background Data	.101	.907	58.2%	66.2%	.624
11-Scale Battery (N=100)	.079	.926	58.0%	57.0%	.575
Background Data	.122	.890	65.0%	60.0%	.625

Eigenvalues and Wilks' Lambda, measures of separation between groups, remain very weak even after the optimum linear combination had been found. The 16PF produced the best classification results, correctly identifying slightly more than two-thirds of both graduates and non-graduates. For the 11-scale battery, the magnitude of the coefficients on the discriminant function suggests that several of the scales are potentially good predictors. The relatively small sample (N=100) may have prevented better classification accuracy.

b. Predicting Graudates' Performance in New Units

Using the original subsets of independent variables, discharge categories were predicted for the 263 graduates to whom the 16PF was administered. Table 2 presents the discriminant functions and the classification results for the long-range prediction problem.

Table 2

Predicting Discharge Categories for Graduates Returned to New Units

16 Personality Factors (N=263)		Background Variables (N=263)	
-.601	H (Socially Bold)	.892	Months' Service Remaining
-.507	Q2 (Independent)	-.363	Offense Category
-.496	N (Astute, Shrewd)	.238	Court Martial Category
-.382	F (Happy-Go-Lucky)	-.219	Age
.354	A (Outgoing, Friendly)	.171	Number of Dependents
-.277	C (Stable Emotionally)	-.166	Marital Status
-.271	G (Persevering)	-.070	Race
-.265	Q4 (Tense, Frustrated)	.031	Education Completed
-.260	O (Apprehensive)	-.024	Highest Pay Grade
.231	Q1 (Experimenting)	.005	GT Score
.203	L (Suspicious)		

	Group Centroids	
0.218	Honorable Discharges	-0.516
-0.433	Other Separations	1.027

Predictors	Eigenvalue	Wilks' Lambda	Honorable Discharges Correctly Identified	Other Status Correctly Identified	Predictive Validity
16PF (N=263)	.095	.913	89.7%	20.5%	.665
Background Data	.534	.651	89.1%	64.8%	.810

The 16PF produced a discriminant function whose largest coefficients describe the false positives--the 88 graduates who failed to earn Honorable Discharges after returning to new duty assignments. They are characterized by the personality inventory as uninhibited (H), independent (Q2), more sophisticated (N), and carefree (F), when compared with their more successful counterparts. Separation between the groups remains quite weak, however, and the 16PF obviously failed to correctly identify those individuals who failed in subsequent duty assignments. The inventory misclassified nearly 80% of these eventual failures as graduates who would eventually earn Honorable Discharges.

In contrast, the 10 background variables produced good separation between the two groups, correctly identified nearly 90% of the Honorable Discharges and over two-thirds of the failures, for a predictive validity of .81. The discriminant function produced with these variables indicates that the amount of time remaining to serve on active duty is clearly the single most important consideration.

Discussion

The purpose in predicting graduation or failure within the training program was to obtain diagnostic information from the discriminant functions, not merely to replicate the decisions of the team cadre. We now know, for example, that graduates tend to be more conforming and more persevering than those individuals who fail to complete the program. This generalization breaks down, however, when we examine success and failure in subsequent assignments. Here, the background variables become far superior predictors of the type of discharges that graduates will eventually receive.

It is possible, if not probable, that Retraining Brigade cadre reinforce conforming behaviors during the short (two-month) program, while denying the individual sufficient opportunities to perform independently of supervision. In other words, the trainee may not experience the kind of "freedom to fail" that he eventually encounters if he is returned to duty. This explanation appears even more logical in view of the fact that many graduates who fail to obtain Honorable Discharges get into trouble after duty hours and/or independently of the normal duty performance requirements. When the graduate is returned to duty, the new freedom may require qualities of self-initiative and self-responsibility which, in many individuals, are lacking.

In April, 1980, the Brigade's Research & Evaluation Division proposed that all candidates for graduation should be screened on the basis of the individual standard score on the discriminant function produced with the 10 background variables. Originally rejected, the concept was later reviewed and endorsed by the Deputy Chief of Staff for Personnel, LTG Maxwell Thurman. By then, validation had been completed with a new sample of over 2,000 graduates returned to duty, utilizing a discriminant function including 12 background variables and offering a predictive validity approaching .85. Since May, 1982, all candidates for new duty assignments have been screened using this model. Within the next two years, after

recent graduates have had sufficient time to complete military service, Honorable Discharge (ETS) rates for graduates returned to duty are expected to reach 82%, a significant improvement over the prevailing rate of about 62% for recent years. The technique also retains the additional advantage of permitting the Retraining Brigade Commander to control both the quantity and quality of graduates returned to duty, consistent with the Army's enlisted strength requirements.

References

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. Handbook for the Sixteen Personality Factor Questionnaire. Champaign, Illinois: Institute for Personality and Ability Testing, 1970.

Edwards, A. L. Manual for the Edwards Personal Preference Schedule. New York: The Psychological Corporation, 1959.

Fox, L. J. Predicting reassignment and subsequent duty performance for 550 Army prisoners: a discriminant model. (Doctoral dissertation, Kansas State University, 1980). Dissertation Abstracts International, 1980, 41, 1951-A.

Georgoulakis, J. M. Unpublished monograph. Fort Riley, Kansas: U.S. Army Retraining Brigade, 1982.

Georgoulakis, J. M., & Fox, L. J. Personality traits of U.S. Army prisoners. Paper presented at the Annual Conference of the Military Testing Association, San Antonio, Texas, November, 1982.

Gough, H. G. Manual for the California Psychological Inventory (Rev. ed.). Palo Alto, California: Consulting Psychologists Press, 1964.

Hudson, W. W. Manual for the Index of Self-Esteem. Princeton, N. J.: Educational Testing Service, 1974.

Nie, N. H., Hull, D. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical Package for the Social Sciences. New York: McGraw Hill, 1975.

Rosenberg, M. Self-Esteem Scale. Princeton, N. J.: Princeton University Press, 1965.

Disclaimer. The opinions expressed in this study represent the views of the authors alone, and do not represent the official position of the U.S. Army Retraining Brigade or the Department of the Army.

Personality Traits of U.S. Army Prisoners

CPT James M. Georgoulakis
Chief
Research & Evaluation Division
U.S. Army Retraining Brigade
Fort Riley, Kansas

Dr. Lawrence J. Fox
Research Psychologist
Research & Evaluation Division
U.S. Army Retraining Brigade
Fort Riley, KS

The study of personality traits among military prisoners is both extensive and somewhat conflicting. An Army Chaplain (Berbiglia, 1971) administered the Taylor-Johnson Temperament Analysis (Taylor, et al, 1968) to confined AWOL offenders in post stockades at Fort Bliss, Texas, and Fort Polk, Louisiana. Results revealed a similiarity in test profiles which was subsequently termed the "AWOL Syndrome." Later tests were conducted with 800 men in a Fort Bliss artillery battalion and with 803 members of an advanced Individual Training Battalion at the same installation. As a result of these studies, Berbiglia concluded that the T-JTA identified individuals with various problems who were not apparent to their company commanders. Further, he reported that AWOL rates were drastically reduced by providing counseling for those men whose test patterns matched the "AWOL Syndrome." However, additional research by Bell, Kristiansen, and Houston (1974) and Frass and Fox (1972), among others, failed to validate the "AWOL Syndrome." Additional research with military (Army/Air Force) prisoners was conducted by Gough and Peterson (1952) utilizing the Socialization Scale of the California Psychological Inventory. Results of the investigation indicated significant differences in the scale between first time offenders and recidivists. However, additional research by Thorne (1963) failed to find such differences. In light of these differences, the following two studies were undertaken at the U.S. Army Retraining Brigade (USARB), Fort Riley, Kansas, to determine if measurable differences exist between prisoners who successfully complete the USARB training program from those who do not. The USARB training program consists of 7 weeks of training designed to place a soldier under sustained physical and mental stress within a stringent military environment. This stress is considered essential to the rehabilitation process.

Methodology

In the first study (Study A), the Sixteen Personality Factor Questionnaire (Cattell, Eber, and Tatsuoka, 1970) was administered to 550 prisoners prior to entering one of the eight training teams. In the second study (Study B), a battery of seven scales from the California Psychological Inventory (Gough, 1957); two scales from the Edwards Personal Preference Schedule (Edwards, 1959); Rosenberg's (1965) Self-Esteem Scale; and Hudson's (1974) Index of Self-Esteem were administered to 260 prisoners prior to entering the USARB program.

In both studies, all the prisoners who were administered the instrument were followed throughout the duration of the program. Upon completion of the program, the prisoners were placed into one of two groups, graduate and nongraduate, depending upon completion/noncompletion of the program. The groups were then randomly reduced to 100 in each group. All data was keypunched and analyzed utilizing the computer program for the *t*-test as contained in the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, and Brent, 1975).

Findings

The results from the Sixteen Personality Factors are shown in Table 1 and the results from the other personality measures are shown in Table 2.

The results of the two studies are consistent and suggest that graduates of the retraining program have more self-control, a better sense of personal responsibility, and are more sociable than those who fail to complete the program. Nongraduates, on the other hand, tend to be more independent, more expedient, careless, indifferent, and agressive. It is important to note that these differences exist *a priori* and are not casual effects of the program.

Discussion

From the results obtained from these two investigations, it appears that individuals who complete the training successfully may well have personalities better suited to the specific requirements of the Retraining Brigade program, and probably to some extent, to the Army environment in general than their nongraduate counterparts. Today, with the Army meeting its recruitment goals with high quality accessions, a case could be made for sending only those prisoners who have the greatest potential for success to the United States Army Retraining Brigade.

TABLE 1

Results from the Sixteen Personality Factor

Factor	Graduates (N=287) $\overline{X}$	(SD)	Nongraduates (N=263) $\overline{X}$	(SD)	F	Sig
A - Reserved/Outgoing	5.69	(1.67)	5.08	(2.04)	14.85	.0001
B - Dull/Bright	4.65	(1.69)	4.73	(1.83)	0.29	
C - Affected by Feelings/ Emotionally Stable	4.97	(1.72)	4.17	(1.95)	25.98	.0000
E - Humble/Assertive	5.13	(2.00)	5.86	(2.02)	18.15	.0000
F - Sober/Happy-Go-Lucky	5.29	(1.85)	5.17	(2.10)	0.54	
G - Expedient/Conscientious	6.22	(1.57)	5.42	(1.74)	31.86	.0000
H - Shy/Venturesome	5.63	(1.85)	5.11	(2.02)	9.56	.002
I - Tough-minded/Tender-minded	6.30	(1.52)	6.40	(1.77)	0.47	
L - Trusting/Suspicious	5.49	(1.88)	5.93	(1.96)	7.19	.007
M - Practical/Imaginative	6.23	(1.92)	6.19	(1.82)	0.05	
N - Forthright/Astute	5.84	(1.82)	6.05	(1.91)	1.77	
O - Self-Assured/Apprehensive	6.03	(1.77)	6.47	(1.93)	7.73	.005
Q_1 - Conservative/Experimenting	5.41	(1.95)	5.54	(1.92)	0.64	
Q_2 - Group Dependent/Self-Sufficient	5.84	(1.64)	6.62	(2.09)	23.44	.0000
Q_3 - Undisciplined Self-Conflict/ Controlled	6.30	(1.74)	5.35	(2.04)	34.68	.0000
Q_4 - Relaxed/Tense	5.72	(1.99)	6.25	(2.01)	9.73	.002

TABLE 2

Results from the Other Personality Measures

	Graduates (N=100)		Nongraduates (N=100)			
	$\overline{X}$	SD	$\overline{X}$	SD	*t*	Sig
1. Index of Self-Esteem (Hudson)	26.07	12.58	28.54	13.47	-1.34	.09
2. Rosenberg's Self-Esteem Scale	23.50	2.54	22.91	2.51	1.65	.05
3. Autonomy Scale (Edwards)	10.40	4.57	11.90	4.51	-2.34	.01
4. Endurance Scale (Edwards)	12.74	3.26	12.55	3.04	0.43	.33
5. Responsibility Scale	21.79	5.25	20.36	5.61	1.86	.03
6. Self-Control Scale	23.64	7.52	21.90	6.98	1.70	.04
7. Social Presence Scale	34.55	5.44	34.48	5.43	.09	.47
8. Well Being Scale	28.96	5.48	27.79	6.43	1.39	.08
9. Dominance Scale	25.71	5.66	24.98	6.15	.87	.19
10. Self-Acceptance Scale	22.02	4.02	21.05	3.74	1.77	.04
11. Socialization Scale	26.90	4.99	24.98	5.38	2.62	.01

REFERENCES

Bell, D.B., Kristiansen, D.M., & Houston, T.J. An evaluation of two system for reducing discipline failure in BCT. ARI Technical Paper ______.

Berbiglia, J.C. The AWOL Syndrome Supplement 3: Taylor-Johnson Temperament Analysis Manual. Los Angeles: Psychological Publications, 1971.

Cattell, R.B., Eber, H.W., & Tatsuoka, M.M. Handbook for the Sixteen Personality Factor Questionnaire. Champaign, Illinois: Institute for Personality and Ability Testing, 1970.

Edwards, A.C. Manual Edwards Personal Preference Schedule. New York: The Psychological Corporation, 1959.

Frass, L.A., & Fox, L.J. The Taylor-Johnson Temperament Analysis "AWOL Syndrome": A Further Evaluation: U.S. Army Correctional Training Facility, Technical Paper, 1972.

Gough, H.G. Manual for the California Psychological Inventory (Rev.Ed.). Palo Alto, California: Consulting Psychologist Press, 1964.

Gough, H.G., & Peterson, D.R. "The identification and measurement of Predispositional Factors in Crime and Delinquency." Journal of Consulting Psychology, 1952, 16, 207-212.

Hudson, W.W. Manual Index of Self-Esteem. Princeton: Educational Testing Service.

Nie, N.H., Hull, D.H., Jenkins, J.G., Steinbrenner, K., & Brent, D.H. Statistical Package for the Social Sciences. New York: McGraw Hill, 1975.

Rosenberg, M. Self-Esteem Scale. Princeton: Princeton University Press, 1965.

Taylor, R.M. et al. Taylor-Johnson Temperament Analysis Manual. Los Angeles: Psychological Publications, 1968.

Thorne, G. "Discriminations Within the Delinquency Continuum Using Gough's Socialization Scale." Journal of Consulting Psychology, 1963, 27, 183.

Disclaimer. The opinions expressed in this study represent the views of the authors alone and do not represent the official position of the U.S. Army Retraining Brigade or the Department of the Army.

ABSTRACT

Vocational Interests/Aptitudes as Predictors of USCG Academy Cadet Performance

Robert L. Frey, Jr.

US Coast Guard Headquarters

Prediction of success versus failure for cadets in the US Coast Guard Academy is a vital concern. For the graduating classes of 1978 through 1981, the attrition rate was approximately 48%. However, academic ability measures alone do not provide sufficient predictive power; the minimum entrance score on the SAT eliminates a high percentage of those applicants who would drop out solely for lack of academic ability. Upon entering the CG Academy, the cadets take the Strong Campbell Interest Inventory (SCII). To improve the prediction process, six "general occupational theme" scores from the SCII were used in combination with SAT scores. General occupational theme scores characterize a person with respect to six idealized occupational interest personality types as described by Holland (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional). Multivariate analysis of variance was used with Academic Major and Graduation/Attrition as the independent factors; the six occupational interest scores and two SAT scores (verbal, math) were the criteria. The relative importance and incremental predictive power of occupational interest dimensions and academic ability as predictors of academic major and graduation were determined.

ABSTRACT

Relative Time Spent Rating Scales: A Historical Perspective

Sharon K. Garcia

Air Force Human Resources Laboratory

Relative time spent rating scales are used as the primary measuring device in task-oriented job inventories. These scales permit incumbents to report on the amount of work time they spend on each task relative to the amount of time they spend on other tasks performed. Measures of relative time spent are currently being collected by the Air Force and other governmental agencies; however, no consensus has been reached regarding the most efficient and accurate scale format to use. This paper tracked the evolution and development of relative time spent rating scales from early attempts at direct estimation of time spent through the use of various relative time spent scales. Research aimed at evaluating and comparing various scale formats was reviewed and conclusions were drawn concerning their usefulness in occupational analysis.

AD P000873

CAREER ATTITUDES OF ROTC CADETS AND COLLEGE STUDENTS

Arthur C. F. Gilbert, Ph.D.
US Army Research Institute for the Behavioral and Social Sciences[1]
Alexandria, Virginia 22333

Lucy B. Wilson, Ph.D.
Booz-Allen & Hamilton, Inc.
Philadelphia, Pennsylvania 19106

The research reported in this paper is part of a larger research program being conducted by the US Army Research Institute for the Behavioral and Social Sciences (ARI) on cadet retention in the Senior Army ROTC Program. In this research variables that influence the cadets' decisions to join, remain in, or leave the program will be identified. Based on the findings of this research, it is anticipated that a number of strategies to enhance retention in the ROTC program can be formulated.

Earlier ARI-sponsored research (Armstrong, Farrell, & Card, 1979) investigated the attitudes and characteristics of ROTC cadets and college students and made comparisons between the two groups on a variety of dimensions. The purpose of this investigation, as in the 1979 effort, was to evaluate and compare the attitudes and values of ROTC cadets and other college students with respect to career aspirations, the Army, and the Army ROTC program. An additional purpose was to contrast these attitudes and values with those of cadets and college students over a three-year period. The information obtained as a result of this research will be combined with other information obtained through literature review and focus group interviews to form the basis for the development of instruments that will be used in the ROTC cadet retention research.

METHOD

A sample of 1,120 students from 11 colleges participated in the research. Selection of the colleges attempted to replicate those used in the 1979 research and accommodated college size (e.g. large, small) and representation by ROTC geographical region. This was accomplished even though 13 colleges were used in the 1979 research.

A slightly modified version of the 1979 questionnaire (Armstrong, Farrell & Lord, 1979), developed for ARI by American Institutes for Research, was used. The questionnaire was divided into four sections that covered

[1]The views expressed in this paper are those of the authors and do not necessarily reflect the views of the US Army Research Institute or the Department of the Army.

(1) background information, (2) school life, (3) career plans, and (4) knowledge about and attitudes toward the military and ROTC. The last section was divided into two parts, one for cadets and the other for non-cadets.

The original questionnaire was updated in two ways. In two media questions regarding magazine readership and radio program preferences, the pre-coded answer categories on the 1979 questionnaire were expanded to incorporate all previously volunteered answers. That is, if students in 1979 reported reading a magazine not then listed, it was included in the modified questionnaire. The second update involved three new items regarding possible changes in the program that might enhance the attractiveness of the ROTC program.

On each of the college campuses, a coordinator for the research was designated from the staff of the ROTC unit. The coordinators made arrangements for questionnaire administration in the ROTC classes and also contacted college instructors to have the questionnaire administered in required freshman or sophomore classes for non-cadets. Completion of the questionnaire required approximately 45 minutes. Participation in the research was voluntary and subjects were not asked to identify themselves.

All questionnaires were returned to a central location. The analyses reported in this paper involved a series of cross-tabulations to determine differences in response patterns between ROTC cadets and college students who were not cadets.

RESULTS AND DISCUSSION

The sample of 1,120 students was predominantly male (66%) and white (68%) as in the 1979 survey. Unlike the previous survey which was almost equally divided between cadets and non-cadets, about 60% of this sample were enrolled in either Military Science I or II. Most of the students were reared in the South in a small town or city. This same pattern occurred in the 1979 survey and is the result of overrepresentation of southern colleges.

Students in the present survey are older by a year than they were in the previous effort, with the ROTC cadets being significantly younger (19.85 years) than the non-ROTC cadets (21.06 years). Mean parental income is reported to be higher now than before, but in line with inflation since the previous survey.

Cadets and non-cadets share the same media habits. They direct their attention mainly to newspapers, general radio, campus newspapers, and TV. ROTC cadets are more likely to read sports and outdoor magazines, while non-ROTC cadets are more likely to read home service and women's magazines. However, this is probably merely a function of the gender composition of the samples. Campus newspapers and radio were included in this survey (and not in the 1979 effort) as potentially useful types of media. The campus newspaper is clearly a popular choice with all students, although campus radio broadcasting receives very little audience support. Students report exposure to numerous magazines and appear to be "reachable" through several general

and focused vehicles. Across the campus, the most popular magazines are the weekly news-oriented ones: Time, Newsweek, and Sports Illustrated. Also widely read are TV Guide, Reader's Digest, U.S. News and World Report, National Geographic, and People. Although ROTC cadets report more exposure to more magazines than non-ROTC cadets, their choices of reading materials do not differ importantly.

The TV preferences of students in many ways parallel those of the American public at large. M*A*S*H is the overwhelming first choice among all groups of students. Other popular choices are the continuing dramatic series of Hill Street Blues, Dynasty, and Dallas. Also popular is 60 Minutes. This pattern is somewhat different than two years ago, when student TV viewing was heavily skewed toward comedy series. These changing patterns are in line with the shifting tastes of the general TV audience. FM programming is a universal favorite among students and will provide the widest reach into the campuses.

Cadets have closer ties to the military and are more knowledgeable about Army life than non-cadets. A finding from the 1979 survey, confirmed in the present study, is that ROTC cadets have more contacts with the military. They more often have good friends and relatives who either were or are ROTC cadets themselves or who have seen military service.

Information about ROTC reaches students through multiple channels--some of which are interpersonal and some media-based. Friends, ROTC personnel on campus, and recruiters all play a role in getting out the message. On the other hand pamphlets, radio/T.V., magazine, and newspaper ads also serve to make students aware of the program. Program awareness and scholarship awareness are not gained concurrently. Students hear about ROTC before becoming aware of scholarships. In fact, it may be because of their awareness and interest in ROTC that they learn about the Scholarship Program. This relationship is demonstrated by the types of information sources used to learn about the Scholarship Program; they are primarily military-related--ROTC personnel on campus, recruiters, and brochures. It is also supported by the fact that one in five non-cadets are totally unaware of ROTC scholarships. As the scholarship is perceived to be an attractive feature of the ROTC program, early and consistent communications about it across all groups will be desirable.

Not surprisingly, ROTC cadets professed more knowledge about ROTC than non-cadets and demonstrate this knowledge. Cadets answer more ROTC/Army knowledge questions correctly. As found in the earlier survey, non-cadets tend to overestimate the obligations of ROTC and underestimate some of the benefits. For example, non-cadets think summer camp is required every year of college but do not recognize that cadets receive a $100 stipend as freshmen and sophomores. The patterns of response to the 1982 and 1979 surveys are remarkably similar. Nearly all respondents know that ROTC is available to men and women and that postgraduate training is available to officers. They consistently err in thinking that all officers are obligated to serve four years of active duty.

As would be expected, cadets find the ROTC program more attractive than non-cadets. However, all students rate highly the guarantee of a job after college and the Scholarship Program. Cadets and non-cadets are consistent in that the requirement for obligated duty after college is valued least by both groups. It should be noted that one feature of the ROTC program, that is, subsequent military service, is perceived as both a plus and a minus. When students think of service as guaranteed employment in this uncertain economy, they find that to be very positive. However, when their attention is focused on the fact that this commits them to a specified period of service, they tend to dislike the obligation. Communications about the ROTC military service requirement need to be particularly sharp when addressing this issue and to convey the opportunities without the perceived liabilities.

Echoing their concerns for employment, students say job security is the most attractive feature of Army life. Officer pay and fringe benefits are also highly rated. Overall, ROTC cadets find the Army more to their liking than non-ROTC cadets. This is shown through higher ratings given to individual features and more aspects of Army life being positively evaluated. Although half of all students would serve in the military if needed, cadets are more likely to perceive it as their duty, whereas most non-cadets have not given military service much thought.

Only about three in ten students had Junior ROTC available to them, and, for the most part, this was an Army program. Only one in ten participated in any Junior program. The attractive and unattractive features of the Junior program parallel those of college ROTC. That is, instructors and the quality of the program are valued, whereas the ROTC cadets and the image of the program are not.

On campuses today, popular college majors are business administration and engineering. The sources of financial aid to college students are multiple, and similarities are found between those used by cadets and non-cadets. The family represents the most important source of money to students. Cadets report ROTC scholarships as an important source, where non-ROTC cadets are more likely to mention other scholarships.

Those closest to the students have the most influence on their educational and career plans. The role model provided by someone in the field is more important to cadets than to non-cadets. This may explain why more cadets have friends and relatives connected to the military and have more contacts and information from ROTC personnel and recruiters.

Cadets have higher salary goals than non-cadets and career choices are congruent with the course of study being pursued in college. Thus, business is a frequent career choice, as is engineering. Cadets, as a group, often seek a career as an Army officer. The ROTC cadets' higher salary expectations may be tied into their views of ROTC and an Army career as a secure position which provides the opportunity for advancement and leadership. On the other hand, it may be that they believe the experience they gain in ROTC and the Army (in addition to their college degree) will contribute to an increased marketability of their skills, should they enter the civilian job market ten years after college.

It is not clear whether students realize that there is opportunity in the Army to pursue activities that draw on their educational training and career interest. It is as if one could not consider a military and technical career at the same time.

Aspects of a job which are highly valued by students include the opportunity to advance, interesting and challenging work, job security, and self-improvement. Essentially, these are the same job factors rated highly in the 1979 survey, only now job security has increased in importance. Cadets also value the chance to be a leader and to be associated with a prestigious organization more than non-cadets. Rating the Army's potential to satisfy various needs along these same job dimensions, it seems that, at least for cadets, the Army can satisfy most of their important criteria. The Army is seen as offering job security, the opportunity to advance and to perform as a leader. In addition, the Army is much more positively rated on most dimensions by cadets than non-cadets, and particularly high ratings are by black cadets. The aspects of the Army which detract from its value in the minds of both cadets and non-cadets are perceived restriction on personal freedom, less opportunity for a stable home life and involvement in the community, and uncertainty in geographic location.

Given that cadets have more friends and relatives with exposure to the military and that the Army is rated highly on many dimensions, it is consistent that cadets think their friends and parents would all rate a military career positively. In general, the cadets are consistent in their positive orientation to the military. They are knowledgeable about and value aspects of a military lifestyle. The dimensions of a job that are important to them are also ones which they think the Army will satisfy. Moreover, the Army is perceived to satisfy many of the aspects which they look for in a job.

Cadets, although aware of and interested in the program by the time they are in high school, tend to delay their decision to join the program until college. This is a departure from the 1979 survey where it was noted that the majority of cadets decided to join ROTC in their high school years. The factors influencing a student to join ROTC are similar to those leading him or her to continue into the Advanced Course--that is, there is support to join from family and friends. Being in the program is consistent with the student's personal system of values and beliefs, and with career objectives. Advertising and information from military personnel do not figure in as factors influencing the decision. It is likely the message that is communicated about the program does not "persuade" anyone to join or continue in the program--rather, it provides information or clarification for students to see how ROTC will meet their personal goals and needs. Slightly less than half of the cadets intend to continue through the Advanced Course, which is about the same as reported in the 1979 survey. Fully one-quarter will not sign up, which again is consistent with the earlier research. The bulk of these who do not intend to continue are female, and a relatively higher proportion are white. It may be that those who joined ROTC found that it did not meet their needs as expected, and therefore they decided not to continue, while those who intend to make the transition believe it will be consistent with their goals.

When four variations on service obligations were linked to the decision to make the transition to the Advanced Course, little impact was noted. Options which offer guaranteed Reserve or National Guard service, a two-year commitment, or a scholarship with an extended or variable tour were presented. The most attractive alternative as measured by the interest shown in it is a two-year service obligation instead of three. About one-third of the cadets state such an alternative would increase their likelihood of continuing in MS III and MS IV. For the most part, the alternatives tested are met with indifference. More than half state the changes would neither increase nor decrease their likelihood of continuing in the program. This reinforces the notion that participation is maintained if it appears to fit one's needs, and if that link cannot be established in the cadet's mind, then the program is abandoned. Cadets are split equally about whether or not to continue ROTC without subsistence. A surprisingly small group of cadets say they would join the Army even if they were not required to do so by contract. As with the previous survey, cadets show a slight inclination toward not joining. For the most part, cadets have not given much thought to their military service. A sizable group are unsure which type of Army service they would prefer and the majority do not know how long they would serve if they joined, and nearly half would not seek a career in the Army. The same optional program changes presented to cadets were evaluated by non-ROTC cadets. In all cases--whether the choice was guaranteed Reserve or National Guard duty, a two-year obligation, or a scholarship with extended or variable tour--more than half of the non-cadets would not be persuaded to join or stay (if they were dropouts) in the Army. Less than one in five would be attracted by any of the proposed alternatives. The students' needs and ROTC or the Army's perceived ability to meet these desires may be the key to attracting and retaining more students. The program changes will give an added appeal but are unlikely to function as inducement if the basic compatibility between needs and satisfaction is not perceived.

REFERENCES

Armstrong, T. R., Farrell, W. S., & Card, J. J. Subgroup differences in military-related perceptions and attitudes: implications for ROTC recruitment. (Research Report 1214). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, October, 1979.

THE MATRIX FORMAT FOR TASK INVENTORIES: A FIRST LOOK

Charles D. Gorman[1]
United States Air Force Academy

Royal Australian Air Force (RAAF) Occupational Analysis is an attempt to determine what people in a specific career area (mustering) do on their jobs. The procedure employed is to provide members of a mustering with a survey instrument (called a task inventory) and ask them two questions. First, they are asked which tasks they perform in their present jobs; and second, they are asked to rate on a scale from 1 to 9 the relative amount of time they spend on each task they perform. From responses to these two questions, it is possible to determine, for any subgroup within the mustering, the percentage of that subgroup who perform each task and the percentage of the subgroup's time which is spent on each task.

Traditionally, task inventories are presented to respondents in linear format as shown in Figure 1. Respondents check the tasks they perform in boxes opposite the task statements, then time-rate the tasks on computer response sheets.

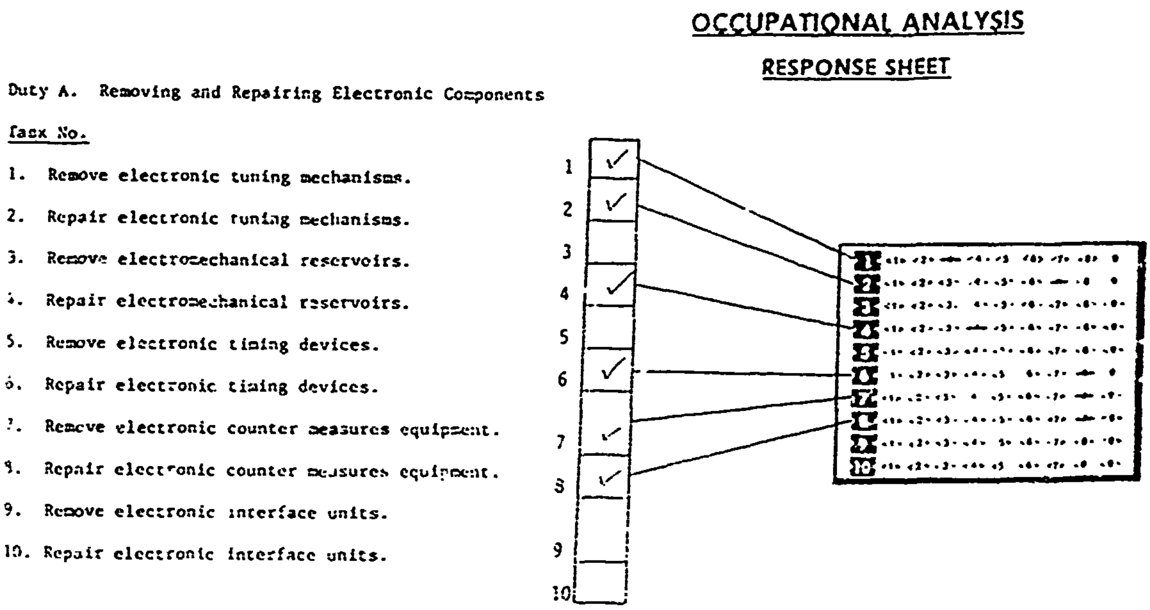


Figure 1. The Linear Format

While the linear format has been successful in several RAAF Occupational Analysis Surveys, it does present problems when a mustering encompasses a particularly large number of tasks. Some technical musterings, for example, involve the potential performance of 3000 or more tasks. Attempts to survey such broad musterings may result in a linear list of tasks which is probably more than the respondent can handle. Even those conscientious enough to try and complete the survey may become alienated through the inventory's sheer length. In either case, the resulting data are of questionable validity and reliability.

This study investigated an alternative format for task inventories administered to broad technical musterings comprising large numbers of tasks. Called the matrix, this format involves identifying pieces of equipment all of which have similar tasks performed on them. An example of the format is shown in Figure 2. Respondents simply check in a matrix of brackets those task-equipment combinations that describe their jobs, then make their time-spent ratings in the appropriate boxes. This format is far less time consuming than the standard linear format and has the added advantage of making it easy for the respondent to spot possible omissions from the task inventory.

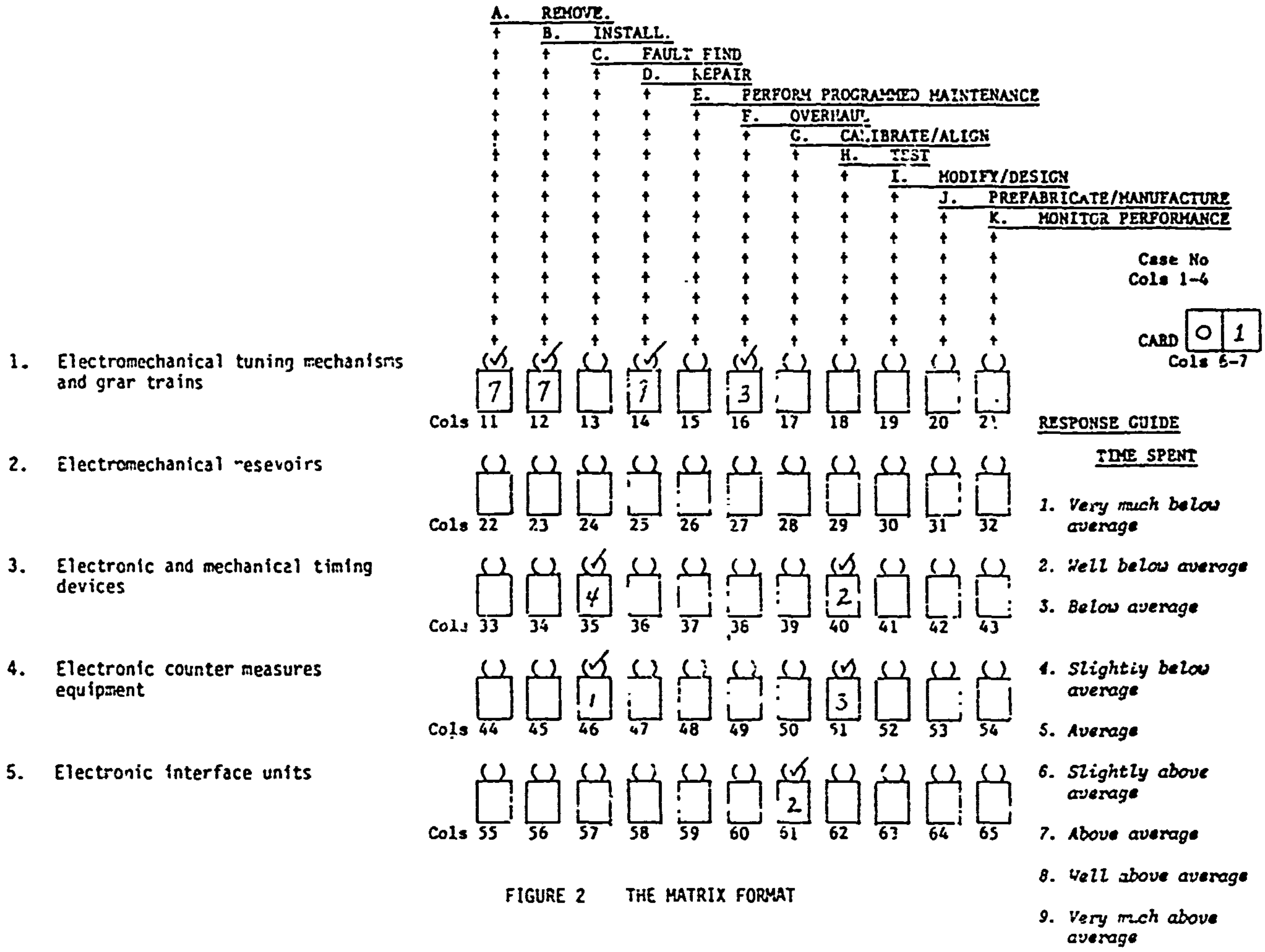


FIGURE 2 THE MATRIX FORMAT

METHOD

The present study was designed to compare data obtained by means of the linear format with data obtained through the use of the matrix format. From a practical perspective, if data obtained by means of the two formats were not significantly different, then one could argue for the use of the matrix format, especially in those circumstances where large numbers of tasks are involved. Conversely, significant differences in data obtained by the two formats would cause one to seriously question the appropriateness of the matrix format for collection of job analysis data.

The procedure for collecting the data for this study was straightforward. Two formats, a linear one and a matrix one, were developed, each containing the same 250 task statements taken from a RAAF technical mustering. Twenty-four technical mustering job incumbents from three RAAF bases served as respondents and completed both formats. Half of the respondents at each location completed the matrix format first, then completed the linear format several days later. For the other half of the respondents at each location, the order of format completion was reversed.

RESULTS

Several comparisons were made between data obtained from the two formats. To gain an overall idea of the similarity of the data from the two formats, two group membership programs were run, one for each format, using the same case as a starter for each. This resulted in two sets of KPATH sequence numbers, one for linear and one for matrix. Case numbers were then placed in order, and the KPATH sequence numbers for the linear format were correlated with the KPATH sequence numbers for the matrix format. The resulting correlation coefficient was .24. When one examined the corresponding diagrams, it was obvious that not only did they differ in appearance, but a given individual was likely to be grouped with different people under the matrix format than under the linear format. In other words, if one were to carry out a job type analysis on the linear diagram, one would obtain different results than if one used the matrix diagram for the job type analysis.

As another overall indicator of similarity of the data obtained under the two formats, a job description was computed for each format, and a group difference description was run. That group difference description showed differences in both percent members performing and percent time spent for 170 of the 250 tasks in the inventory. While the differences for the most part were not particularly large, they were nevertheless present.

To look at the effect of the formats on individuals' responses, job descriptions were computed for each respondent on each format. Then each respondent's time spent ratings under the linear format were correlated with his time spent ratings under the matrix format. Those correlation coefficients are presented by respondent in Table 1. It can be seen that some respondents' ratings appeared relatively unaffected by the format, while others were obviously severely affected. The overall correlation of .86, when composite time spent ratings from all respondents were correlated across formats, was encouraging.

In an effort to determine if percent members performing data were less affected by format than were percent time spent data, diagrams for each format were run based solely on percent members performing data. Results were encouraging. The correlation coefficient between the KPATH sequence numbers obtained from the two formats when only percent members performing data were considered was .65. While the diagram based on the linear format was a slightly different shape from the diagram based on the matrix format, there was a greater tendancy for specific individuals to be grouped similarly in both diagrams than was the case in the percent time spent diagrams. In other words, job type analyses would produce more similar results regardless of format if only percent time spent data were used in the analysis. This is not to suggest that one should use only percent members performing data to conduct job type

analyses. However, it does suggest that one might place more confidence in percent members performing data than in percent time spent data when the data are collected with the matrix format.

TABLE 1.

Correlations of Time Spent Ratings
Linear vs. Matrix Format

Respondent	Correlation Coefficient
001	.70
002	.88
003	.69
005	.62
006	.51
007	.78
008	.16
009	.06
010	.79
011	.56
013	.75
014	.07
015	.90
016	.33
017	.75
018	.49
019	.56
020	.64
021	.77
022	.57
023	.80
024	.44
025	.56
026	.87
ALL	.86

One final comparison was made between the two formats. The mean number of tasks performed under the two formats was computed. Members performed an average of 50 tasks, as reported in the linear format; but only performed an average of 39 tasks as reported in the matrix format. This difference in average number of tasks performed was significant at the .05 level ($t=2.51$, $df=23$ for matched group t-test for the difference between means).

DISCUSSION

The results of this study are equivocal. In some analyses, the matrix format apparently distorted data; in other analyses, data collected with the two formats correlated highly. Specifically, it was generally the case that abberations, apparently caused by differences in the formats used to collect data, were more severe in the case of percent time spent data than in the case of percent members performing data. There are at least two possible explanations for this tendancy, one statistical and one more psychological.

From a statistical perspective, most of the results that throw a negative light on the matrix format could possibly be attributed to the small sample size employed in this study. That is, because of the small sample size, a few divergent raters may have had a disproportionately large effect on the time-spent results of the study. Any future research on the matrix format would want to employ a much larger sample to counteract this effect.

From a more psychological perspective, the matrix format may itself be the cause of disparities within time spent data. When making their relative time spent ratings on tasks they have checked, respondents are instructed to use as their frame of reference their current job in its entirety. If one assumes that the sample format in Figure 3 represents the entire inventory, then the appropriate frame of reference is all tasks checked within the solid line. The matrix format may induce respondents to develop sub-frames of reference around, for example, each piece of equipment or each type of task as indicated by the areas surrounded by broken lines. Use of such sub-frames of reference would completely destroy the meaning of time spent data as it is presently defined and would make interpretation of these data impossible. The fact that much of the disagreement in the data collected between formats is in time spent data could be interpreted as support for this shifting frame of reference hypothesis.

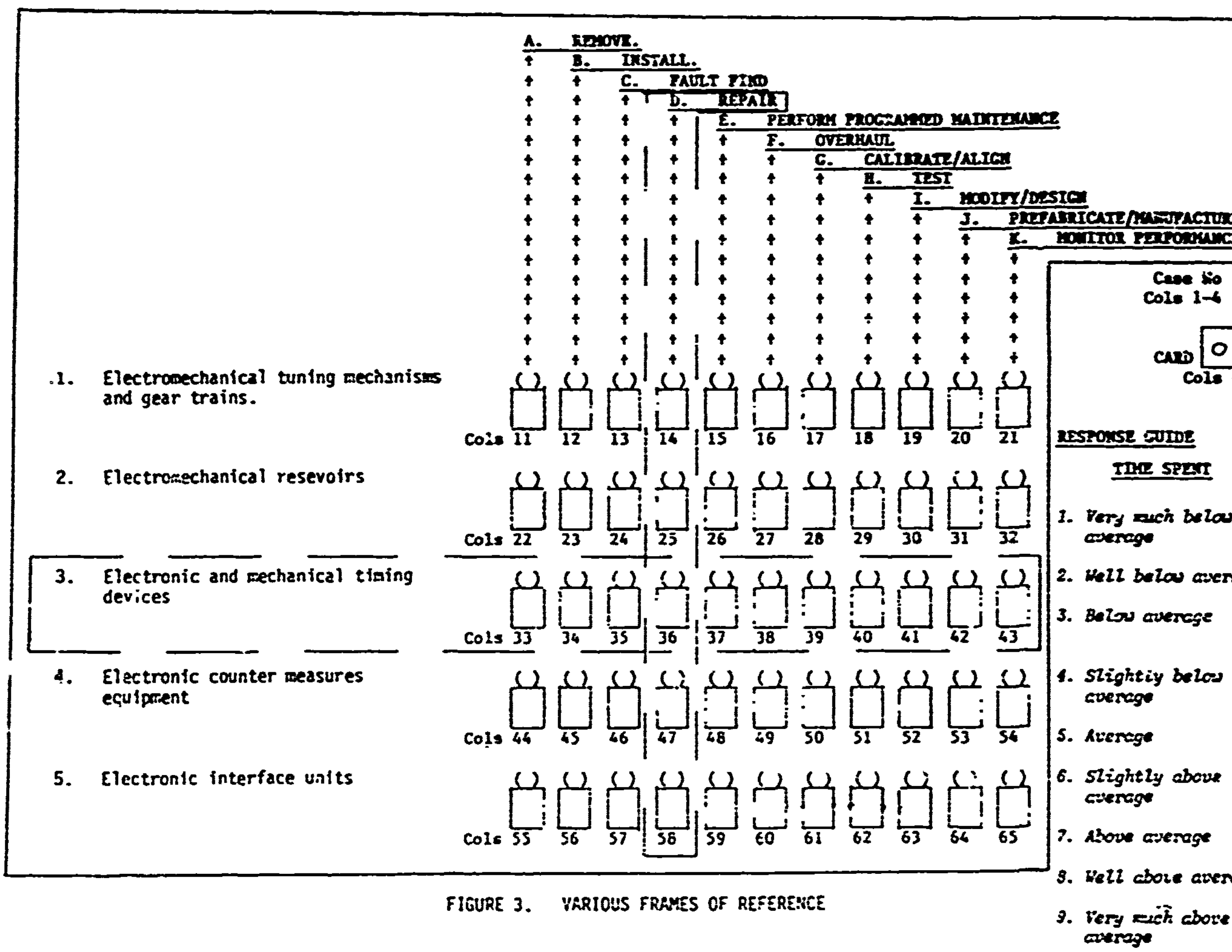


FIGURE 3. VARIOUS FRAMES OF REFERENCE

CONCLUSION

The results of this study indicate that, from a practical standpoint, further research of the matrix format may be fruitful, especially with respect to the collection of percent members performing data. From a more theoretical perspective, investigation of the shifting frame of reference question is certainly an interesting research possibility. Use of the matrix format in constructing task inventories has been shown to be beneficial because of the ease with which subject matter specialists can identify missing tasks. However, use of the matrix format for the collection of job analysis data might best wait until further research is completed.

[1]The opinions expressed herein are those of the author and do not reflect official policy of the USAF or DoD.

STABILIZED GUNNERY TRAINING TECHNIQUES

James H. Harris
(HumRRO)

Stephen L. Goldberg
John E. Morrison
(ARI)

Modern main battle tanks can be fired while on the move. To realize the full potential of these armor systems, tank gunners must be taught to fire from a moving platform. Live-fire exercises, however, are prohibitively expensive in terms of fuel and ammunition costs. This paper presents a program designed to train stabilized gunnery skills in the conduct of fire phase of the Armor OSUT M60A3 program without requiring live-fire exercises. The paper also presents conclusions based on tryouts of two products of the training program.. The training approach has applications to unit training and training on other stabilized tanks (i.e., the M60A1AOS and the M1). The M60A3 tank was chosen over the M60A1AOS because it has the more sophisticated stabilization system and more closely resembles the fire control system on the M1 which had not entered the OSUT inventory when the project began.

The training program content was derived from literature on stabilized gunnery, a hands-on orientation to M60A3 stabilized gunnery, and interviews of subject matter experts. The program material consisted of three products: (1) a knowledge videotape for presenting information on stabilized gunnery, (2) an inexpensive training device for practicing the timing skills of stabilized gunnery, and (3) hands-on exercises for practicing on actual M60A3 equipment skills learned from the knowledge videotape and training device.

Determine Program Content

A review of relevant field manuals and technical manuals indicated two principles that must be followed when firing on the move:

1. Treat each round as a separate engagement. When firing on the move, particularly against moving targets, the rapidly changing tank-to-target relationship makes BOT difficult, if not impossible, to use.

2. Fire only when the gun tube is over the front or rear fenders. The smaller the acute angle between the gun and the line of tank travel, the better the stabilization. Therefore, fire over the flank only as a last resort.

The hands-on orientation began with a review of the "arrangement" of both the gunner and tank commander stations, to include the operation of the fire control system. Then, dry fire engagements were run at various speeds over progressively rougher terrain. In addition to clarifying the mechanics and operation of stabilization on an M60A3, the orientation clarified vividly the major difference between firing from a stationary tank and firing from a moving tank.

The stabilization system of the M60A3 tank is designed to keep the gun tube and sights at the same elevation and direction regardless of the up-and-down or side-to-side movement of the tank. Thus, stabilization aids the gunner in keeping the reticle on target. Nevertheless, there are "error" inputs into this man-machine system which tend to draw the target off the reticle cross hairs, inducing apparent reticle movement with respect to target scene. A primary source of error input, common to moving platform and stationary gunnery, is movement of target relative to firing tank. The critical difference between the two gunnery modes is that, in moving platform gunnery, apparent reticle movement can also be caused by movement of the firing tank. Fortunately, these error sources are somewhat predictable and can be corrected by adjustments in tracking.

Two other error inputs are caused by limitations of the stabilization system itself. The first error source is due to tank movements too large or too fast for the stabilization system to compensate. The second is caused by the linkage of the gun and the sight: If the linkage has some play in it, the sights will appear to jiggle.[1] These errors also induce apparent reticle movement. However, both errors are too fast and unpredictable to be corrected by tracking adjustments. Experienced M60A3 gunners report that to overcome the seemingly random sight movement, that is, the movement of the target relative to the firing tank, the gunner must be able to time his shot because the cross hairs are on the target only momentarily. He must anticipate when the target will approach the center of the reticle and lase and fire prior to its reaching that point. This timing skill is a gunnery component peculiar to firing on the move.

The interviews with subject matter experts were informal and open-ended. A soldier's response to a particular question led naturally to other questions. Some of the information gathered from these interview sessions proved useful during the development phase of the project. Following are some of the questions whose answers helped determine the program content:

1. When firing the M60A3, what is harder about firing from a moving platform (at least the first few times) than firing from a stationary platform?

   Answers:

   a. Timing "pattern"[2] about the target.
   b. Changes in speed of apparent reticle movement when firing tank changes speed.

[1]Discussions with TRADOC Systems Manager (TSM) personnel indicated that much of the "sight jiggle" in early production M60A3 tanks was due to a faulty gun/sight linkage. Mechanical improvements to the older sights have minimized the problem, however.

[2]These "patterns" are the seemingly random reticle movements caused by the three types of error inputs inherent in moving platform gunnery.

2. What do you do to compensate?

Answers:

a. Time shot. This timing, or anticipating, skill is a gunnery component peculiar to firing on the move.
b. Learn to recognize drift patterns and fire on first return to target.
c. Ambush the target.
d. Fire lots of rounds.
e. Let stabilization system operate around target area; gunner just track target.
f. Know speed at which stabilization system smooths out.

Information gathered from the reviews, orientation, and interviews was consolidated and the following principles of firing on the move emerged:

1. Treat each round as a separate engagement.
2. Know the "sweet spot" for your tank.[1]
3. Know reticle drift pattern for your tank.
4. Anticipate "pattern" of reticle movement.
5. Anticipate movement of tank.
6. Fire between front or rear fenders.
7. Fire over flank only as last resort.
8. Press head into browpad, back against seat back.
9. Allow stabilization system to do its work.
10. Lase and lead with either thumb switch.
11. Know that when turret is in STAB mode, don't have to squeeze palm switches to traverse or elevate and depress turret.
12. Know there is no such thing as a "perfect" sight picture.
13. Know that main gun, within limits, maintains fixed orientation in space regardless of vehicle motion.
14. Take up same sight picture.

The development of a training program centered around these principles was undertaken. But since the program was to be used during the conduct of fire phase of M60A3 OSUT, certain constraints had to be considered: the relative inexperience of the soldiers; the limits on available time; and, a scarcity of tanks, main gun ammunition, gasoline, and ranges suitable for moving tank gunnery. Thus both the analytically derived gunnery principles and the prevailing program constraints guided the design of training materials.

---

[1]The "sweet spot" speed is the speed where the apparent reticle movement is minimal. The sweet spot differs for each tank depending on such factors as terrain type.

Develop Training Materials

The developmental approach to training was straightforward: provide performance-oriented instructional events in which the soldier could acquire (a) knowledge of the relevant stabilized gunnery principles, and (b) skill in their application. Too, the approach called for a training medium that was inexpensive yet permitted a level of visual realism sufficient to display realistic stabilized reticle movement in relation to recognizable targets. A video display linked to a simple response mechanism met these requirements.

Tank targets at various speeds and ranges were filmed through the stabilized sight of an M60A3 moving tank. Films of these targets were sorted out on the basis of clarity and demonstration of the stabilized gunnery principles; then, arranged in terms of engagement difficulty. Two videotapes, one for training knowledge of stabilized gunnery principles, the other for practicing those principles were prepared. After the videotapes were prepared, a series of exercises was developed to enable soldiers to practice on M60A3 tanks what they had learned on the videotapes. The exercises are designed to be used anytime the soldier is in the gunner's seat and the tank is moving.

The knowledge videotape (KT) presents the firing on the move principles in terms of their knowledge components. The practice videotape (PT), when coupled with a simple response device, enables practice of some skill components of the firing on the move principles. In general, the videotapes are to be used during training to:

1. Familiarize soldiers with the "patterns" of reticle movement about the aim point during stabilized gunnery engagements. (KT)
2. Demonstrate the correct point in the "pattern" to lase and fire. (KT)
3. Provide practice in "anticipating" the reticle movement about the aim point during stabilized gunnery engagements. (PT)
4. Provide practice in lasing and firing. (PT)

The knowledge videotape presents twelve situations in increasing order of engagement difficulty. Engagement difficulty is presumed to increase as range to target increases and firing tank speed, target speed, or both increase. The M60A3 orientation focused the scope of the training content on target engagements where the firing tank is traveling at speeds of 10 MPH or less, the target tank is stationary or traveling at 10 MPH, and the firing tank-to-target range is 1600 meters or less. The 12 situations are followed by five new situations in which the correct lase and fire points during the reticle movement are demonstrated. In addition, on the last two situations, the correct technique for adjusting fire is discussed and demonstrated. Narration describing the firing on the move principles as they are presented is provided throughout the videotape.

A series of five exercises was developed to enable soldiers to practice on M60A3 tanks some of the things presented in the knowledge videotape and practiced using the device and practice videotape. The exercises comprise the essential requirements for acquiring proficiency in moving platform gunnery on the M60A3 tank. They should be practiced whenever possible. The practice can be done formally, during scheduled training time, or informally, whenever the tank is moving and the soldier is in the gunner's position.

Exercises were developed to include:

Exercise 1: Taking up the correct position in the gunner's seat.
Exercise 2: Determining the sweet spot for the tank on which he is the gunner.
Exercise 3: Tracking targets when the tank is moving.
Exercise 4: Lasing and firing on targets when the tank is moving.
Exercise 5: Reengaging to adjust fire.

The videotape training products were tried out, revised based on the tryout results, and tried out a second time. The data obtained in the two tryouts and the constraints which guided the design of training materials permitted the following conclusions:

- The stabilized gunnery knowledge videotape is an effective procedure by which to present information on moving platform gunnery to soldiers. They expressed positive attitudes towards its use in a training program. The KT can be group administered using equipment available in any OSUT battalion.
- Soldiers indicated that the stabilized gunnery practice tape device enabled them to gain confidence both in their ability to anticipate the apparent reticle movement and to respond to the movement. The PTD is relatively inexpensive to produce and can be set up in a dayroom or corner of a classroom.
- The stabilized gunnery practice tape device is of little value in training soldiers to perform the tracking element of moving platform gunnery. Ss did, however, tend to decrease their lasing and firing time across sets of engagements, although with one exception these time improvements were not significant. These empirical results seem to back up Ss feeling of confidence gain.

The principles presented on the videotape are:

- Three contact points
  - Press head firmly against browpad.
  - Press lower back against gunner's seat backrest.
  - Place feet flat on turret floor.
- Reticle movement
  - Movement caused by stabilization system.
  - Influenced by speed of tank and type of terrain.
  - The speed where vibration in sight picture smooths out and reticle jumps around less is the "sweet spot."
- Tracking
  - Let stabilization system make fine corrections around the target area.
  - Use gunner's control handles to track the target.
- Front deck
  - Lase and fire only when gun tube is over the front deck, unless . . .
  - You encounter a surprise target on your flank.
- Lase and fire
  - Depress and hold either palm switch.
  - Track for at least 1-1/2 seconds.
  - Anticipate reticle movement toward center of mass.
  - Lase immediately when it moves toward center of mass.
  - Fire immediately when reticle moves again toward center of mass.
- Adjust fire
  - Reengage technique to adjust fire.
  - Release and then depress gunner's palm switch.
  - Track target.
  - Relase.
  - Fire a second round.

The practice videotape presents 18 situations of 20 seconds each. The first nine situations are presented in increasing order of difficulty; then, the same nine situations are presented in random order. The videotape is to be used with a very simple mechanical response device called the Practice Tape Device (PTD) which includes a set of M60A3 gunner handles and periscope. The gunner handles are not responsive; the device provides practice only on timing (anticipating) not tracking. The device is designed so that the soldier observes the video display through the periscope and lases and fires when he determines the sight picture to be correct for lasing and firing. When the soldier thinks the sight picture is correct for lasing, he presses either gunner's thumb switch to set lead and fire the laser. The videotape "freezes" and the accuracy of his response, in terms of deflection (left or right) and elevation (short or over), as well as the time to respond can be recorded and evaluated. The device is reactivated after the lasing response is recorded and the soldier presses either firing trigger when the sight picture is correct for firing. Again, the videotape "freezes" and the accuracy of his response as well as the time to respond can be recorded and evaluated.

AD P000876

Assessment of the Hometown
Recruiter Assistance Program

Allyn Hertzbach, Timothy W. Elig, Paul A. Gade, Guy L. Siebold,
Newell K. Eaton, John W. O'Hara

US Army Research Institute
for the Behavioral and Social Sciences

Introduction

The purpose of this paper is to present an assessment of the Hometown Recruiter Assistance Program (HRAP), which is documented in US Army Recruiting Command Regulation No. 601-64, 1981. It describes the results of the investigation of the nomination, selection, and training of recruiter aides.

Background

The HRAP is a tri-service program that returns young military personnel to their hometowns to assist recruiters in a local recruiting station. Recruiter aides, as the Army's HRAP participants are called, come from Training and Doctrine Command (TRADOC) and Forces Command (FORSCOM) installations. Usually, TRADOC aides are sent after completing Advanced Individual Training (AIT); occasionally, aides may be deployed following Initial Entry Training (IET). FORSCOM recruiter aides are selected from regular duty units. All aides are nominated by their enlisting recruiters and approved by their AIT or duty units. The aides are volunteers and usually serve for 45 days on temporary duty (TDY). Their function is to bring in qualified applicants to meet recruiters rather than to recruit.

Evaluating the productivity of aides is difficult because there is no existing basis to fully rate their performance. Aides are credited for individuals they brought to the recruiter who subsequently enlist, but the total effect of the aide's efforts is more subtle than the sum of their recruits. For example, aides can "plant seeds" or lay the groundwork months in advance of an enlistment decision and receive no credit for an enlistment that occurs months after his or her departure. Also, the criteria for receiving credit are not standardized. Some aides might be given credit for the enlistment of an individual whom they did not initially bring to the station, but helped "sell," while others may get no such credit. (Much might depend upon the recruiter's feelings about the recruiter aide.) Finally, there are a myriad of criteria that could be used to evaluate aides, aside from enlistments, which would credit the aide's skill and effort, such as the number of appointments made for the recruiter, number of prospects seen, and level of effort as noted by the station commander or recruiter. Since the job of an aide is getting qualified people to the recruiter, perhaps the ability to bring in interested and qualified people is a better measure of aide performance than the total number of enlistments.

The views expressed in this paper are those of the authors and do not necessarily reflect the views of the US Army Research Institute or the Department of the Army.

Despite the difficulty in measuring a recruiter aide's productivity, there is some evidence that the contribution of recruiter aides is significant. Trautwein and Toomepuu (Note 1) found that recruiter aides made a positive contribution in the recruitment of high school diploma graduates in Armed Forces Qualification Test (AFQT) categories I through IIIA. In this analysis recruiters produced an average of 3.5 of these recruits per quarter; aides contributed .5 of these recruits.

However, a more comprehensive measure of aide productivity is required before the program can be accurately evaluated. Productivity figures do not adequately differentiate among organizations that effectively use aides and those that do not. It is possible that if aides were employed to maximal advantage throughout the US Army Recruiting Command (USAREC), dramatic positive effects could be achieved.

Approach and Method

As previously mentioned, this research paper describes the results of the investigation of the nomination, selection, and training of recruiter aides. Information was collected from the personnel most familiar with the day to day performance of recruiter aides. Station commanders, recruiters. and where possible, recruiter aides were surveyed and interviewed between August and October 1981.

Surveys and Structured Interviews.

The survey consisted of a paper and pencil questionnaire that solicited information about demographics, recruiter productivity, job satisfaction, personality characteristics, and job preferences. The structured interviews covered several topics and the questions were identical for recruiters (RCs) and station commanders (SCs). Responses were usually open ended. Interviews lasted between 1 and 2 hours per person.

Survey and Interview Samples.

Recruiters and station commanders were sampled equally from each of the 5 recruiting region commands. Within, each region, 5 district recruiting commands (DRCs) were selected at random; then 2 recruiting stations were selected from each of these designated DRCs. Due to problems with sample stations an additional 3 stations were visited. The sample included 53 station commanders, 103 recruiters, and 20 recruiter aides. Five ARI researchers conducted the interviews, with each collecting data at different sites. Interviews were conducted in a private location within the station (during normal duty hours).

Nomination and Selection of Recruiter Aides

Station commanders and recruiters were asked about their HRAP nomination practices, who they thought should select recruiter aides, and what qualifying criteria should be met by young soldiers returned to their hometowns to assist recruiters. Respondents also provided estimates of the percentages of their recruits that they nominated for the HRAP. They then estimated the percentage of their nominations that have been returned for duty as recruiter aides. Nominations for the HRAP range from 0 to 100% of recruits. Thirty-six percent of

SCs and RCs nominated between 0% and 10% of their recruits. Forty percent of all respondents nominated between 11% and 50% of their recruits; the remaining 24% of respondents nominated between 51% and 100% of their recruits, a total of 51 SCs and 92 RCs comprised the total of respondents. The pattern of nominations was similar for recruiters and station commanders; and, despite a wide range in nomination rate, most respondents actively nominate aides.

The pattern that emerges from the above analysis is not particularly revealing. Some respondents nominate very few of their recruits while others nominate a majority of their recruits. The clue to nomination practices could lie in a number of possible explanations. But for this discussion, the key question is, how effectiv is the nomination process? Are a reasonable number of aide nominations made and returned to the stations? The first question about nominating practices suggests that SCs and RCs are not reluctant to nominate recruits for the HRAP, though some appear to be more discriminating than others.

The next important consideration is the rate at which aide nominations are returned to the station. Nearly 75% of the respondents reported a return rate of 5% or less. An additional 14% of the respondents reported a return rate of fewer than 33% of their aide nominations. Four percent report a return rate of more than 50% of their nominations; however, these individuals are usually relatively new at recruiting and have made only 2 or 3 nominations and gotten one or two returned. The overwhelming number of respondents get a very low rate of return of their aide nominations.

This finding is supported in general comments or asides that respondents made during the interview. Often, complaints were made that individuals returned were not nominated and/or not qualified by the nominal requirements in USAREC Reg. 601-64. Nearly 30% of the recruiters reported dissatisfaction with aides returned or with the effectiveness of the selection process.

Other evidence describing the view that RCs do not have adequate control of the selection process is found in responses to the question, "Who should select recruiter aides?" Overwhelmingly, SCs and RCs declared that recruiters should be at least part of the selection process and given a powerful voice in aide selection. Eighty-seven percent of all respondents named recruiters solely or in combination with duty or training unit cadre as the individuals who should select recruiter aides. Thirteen percent of the respondents named the training or duty unit only or a board of varying composition.

In addition to the call for increased recruiter control in the selection of aides, respondents detailed a comprehensive list of criteria for determining qualification for selection as a recruiter aide. The thirteen most frequently mentioned criteria are enumerated in Table 1.

Perhaps, one important way to improve the current system would be to make the nomination for the HRAP more discriminating by providing recruiters with a list of criteria and ask that they justify each nomination.

TABLE 1

Recommended Aide Selection Criteria by SCs and RCs

Objective Criteria	SC* %	RC* %
1. High School Degree Graduate	16	30
2. AFQT Category IIIA (or higher)	20	15
3. Delayed Entry Program Performance	26	28
4. Training/Duty Performance	28	21
5. Good Military Appearance	51	34
Subjective Criteria		
1. Popular	22	11
2. Good Attitude & Character	22	11
3. Can communicate	29	28
4. Motivated	15	--
5. Gregarious	66	40
6. Sensible/Smart	12	25
7. Positive Attitude toward Army	52	38
8. Desire to be an aide	32	40

N=53 SCs N=98 RCs

*Percentages do not add to 100%, as respondents often suggested more than one criterion. Each category, however, reflects a respondent only once.

Recruiter Aide Training

Respondents' perceptions of recruiter aide training form the basis for this section. RCs and SCs were asked, "What, if any, training problems exist in the aide program?" The views expressed suggested that current training is inadequate.

Most respondents identified training problems. When RCs and SCs were asked if they thought there were aide training problems, 56% said yes and 26% said no. An additional 18% (of all recruiters) expressed no opinion. Several kinds of problems were identified by respondents, the largest being the inadequacy of training prior to the aide's arrival at the recruiting station (43% of SCs and 59% of RCs).

When asked to enumerate the problems, respondents who had previously stated that there were none were usually consistent and either stated that there were no problems or made no response. Several other SCs and RCs followed up their no problems response by saying that there was virtually no training prior to the aide's arrival at the station because there was too little time and/or money for the provision of training.

The opinion that RCs and SCs expressed about the lack of adequate training for aides prior to their arrival at the recruiting station is supported by the recruiter aides interviewed. Of the 20 aides interviewed, 17 reported fewer than 2 hours of briefing or training prior to being sent to the recruiting station. Often the briefing concerned administrative matters only, i.e., how to fill out forms. There was very little training in the activities that the aide would need to be successful in assisting recruiters. Twelve of the aides requested additional training; they most often desired training that would help them attract qualified individuals to the station. Aides most often desired training in the use of the telephone and in prospecting. They also felt that they needed more product knowledge to successfully perform in the field.

What emerges from these responses is the view that there is little or no training of recruiter aides prior to their arriving at the recruiting station. Even some respondents who do not label this deficit in training a problem are aware of it. Of course, some other respondents may feel that the station can adequately provide the training, and that there are no problems.

Then, respondents were asked whom they thought should train recruiter aides; a majority of respondents (51%) felt that the recruiting station should do the aide training. Fifty-five percent of the respondents who expressed the view that the station should provide training did not mention another command level; forty-five percent of respondents who mentioned that training should be provided by the station felt that other command levels should provide training, as well. The most frequently mentioned command levels were the DRC and/or the area (39%).

Of the remaining respondents, 44% named one or more command levels other than the recruiting station to provide recruiter aide training. The most frequently named single command level was USAREC (12% of the total sample). The DRC in conjunction with USAREC (10%) and Area (11%) was the next most frequently mentioned command for assuming aide training responsibility.

Additional evidence for the need of higher command assistance in the training of recruiter aides was found in examining the curriculum recommended by SCs and RCs (see Table 2). This list of topics requires a sound instructional design, which station personnel could help develop. However, RCs and SCs lack the time and training to design and implement what would be a relatively sophisticated training program. Future efforts need to be directed at developing and testing alternative training programs, in order to identify the most effective and efficient training to be provided.

TABLE 2

Recommended Recruiter Aide Training

	SCs		RCs		TOTAL	
	f	%	f	%	f	%
<u>Knowledge</u>						
Product knowledge	12	24	26	27	30	26
Prequalification & Eligibility	16	31	35	35	51	35
<u>Skills</u>						
Prospecting	17	33	14	15	31	21
Interpersonal/Social	9	18	9	9	18	12
Persuasiveness/Sales	8	16	34	35	42	28
Telephone	19	37	24	25	43	29
General Skills	26	51	70	73	96	65
<u>Conduct</u>	5	10	4	4	9	6
<u>Other</u>	9	18	8	8	17	12
	121*		224*		345*	

N= 96 RCs
N= 51 SCs

147

* Respondents often made more than 1 response so that column percentages add to more than 100%.

In order to develop an effective training program a curriculum and a method of delivering training need to be selected. The results (Table 2) of this research effort could be used as the basis for a training curriculum, although final approval should rest with representative samples of SCs and RCs. Next a field test should compare efficient ways of delivering training to recruiter aides. Sharing of training between the DRC (or Area) and the station could be compared with the station alone providing training. A final decision could be made on the basis of immediate outcomes from the training and later recruiter aide productivity.

Reference Note

Trautwein, M. and Toomepuu, J. Analysis of the Contribution of Recruiter Aides to Recruiter Mission Accomplishment. Program Analysis and Evaluation Directorate. US Army Recruiting Command, Ft. Sheridan, IL. July, 1981.

ABSTRACT

Current Status of Counter-Attrition Programs in the Armed Services

Jack M. Hicks

US Army Research Institute

This was a progress report on the current status of three categories of programmatic effort which show promise for countering first-term enlisted attrition. The program areas are (a) preenlistment education and training, (b) realistic expectations intervention, and (c) correctional retraining.

Computer Assisted Training for Letter Sorting Machine Operators

Joseph M. Hillery, James E. Mahoney
and Timothy J. Bohen
U.S. Postal Service, Washington, D.C.

The U.S. Postal Service has conducted a series of studies on the feasibility of using Computer-Assisted Instruction (CAI) for training of operators of machines using various keyboards. Earlier studies clearly demonstrated a strong possibility of cost effectiveness for CAI training but the exact combination of hardware and software that would show a distinct advantage over the existing training system proved elusive.

The earlier CAI studies in the series used minicomputer hardware. With the development of the microcomputer in the late 1970's, a training package was developed taking advantage of the new developments in microcomputer hardware. By late 1979 a prototype of the system was operational and a pilot test began in early 1980.

The target job for the test of CAI training was that of Multi-position letter sorting machine (MPLSM) operator. The majority of the mail which flows through the distribution network of the Postal Service is sorted by the MPLSM's. Each MPLSM consists of 12 operator consoles from which the mail is sorted into 277 bins. The operators rotate between the tasks of keying mail, keeping the console ledges loaded with mail, and sweeping the mail out of the bins. The vast majority of time and the task of greatest importance is the keying of codes.

The MPLSM operator must key a code for letters presented at the machine paced rate of 50-60 letters per minute, depending on the type of mail to be sorted. The keyboard consists of ten regular piano type keys plus ten prime or piggy-back keys located directly over the others. The letters stop momentarily in the viewing position above the console keyboard to allow the operator to read the address. As the letter starts to move away from the viewing position, the operator keys the appropriate code. At present there are over 37,000 MPLSM operators in 224 Post Offices and the projected yearly need is for an additional 6000 operators.

Method

Participants. All individuals in the study were selected for training from the MPLSM operator selection register in the North Suburban Mail Processing facility, outside Chicago, between the dates of April 18, 1980, and August 9, 1980, in order of their test scores. A few of these people had some previous training on the MPLSM or were current Postal Service employees. Excluding these, 274 people started dexterity training as study participants in the field test.

Information on the trainees was obtained from the employment application and from information available through the North Suburban Personnel Office. The composition of the participants was as follows: age, 17 to 64 with a mean of 27.86; test score range (without veteran preference points), 79 to 100 with a mean of 89.23; and sex composition of 51% females and 49% males.

MICRO (CAI) training. The basic hardware requirements for the training system were completely independent training units with video display, simple graphics capability, relatively high external data transfer rate, hard copy capability, and the ability to accept external inputs. The search for an off-the-shelf microcomputer that best met these requirements resulted in the selection of the Tandy Corporation TRS-80 Model I.

For the field test, the total training configuration consisted of ten training systems and one management system used by the instructor. Each training system included one 16 KV RAM CPU/keyboard combination, 12" B/W video monitor, expansion interface with 32 K RAM, 5¼" mini disk drive, Quick printer II, interface box and MPLSM keyboard. The management system included the same CPU/keyboard, monitor and expansion interface along with three 5¼" mini disk drives and a Microtek MT-80 printer. The interface box used with the training systems was designed so that a variety of keyboards, not just the MPLSM, could be used with the training system.

All the applications programs were developed in-house and all programming was done in TRS-80 BASIC. Some second source utility software was purchased to provide high speed scrolling, sorting, and for external keyboard input. Since one of the most important elements in the training is the development of a sense of timing, a letter was simulated and moved across the CRT screen at predetermined speeds. The MICRO trainer created the image of a letter by using two parallel horizontal lines, a three line address and simulated stamp.

Two aspects of the training lessons made possible by CAI were automatic speed increments and automatic lesson termination. In two of the training lessons, the lesson started at a speed of 46 letters per minute. When the trainee reached a predetermined accuracy level at that speed, the speed was automatically advanced two letters per minute. When the prescribed accuracy level was met at a speed of 60 letters per minute, the lesson was completed and the next lesson was presented.

An additional feature of the CAI training was that of automatic lesson termination. This feature applied to any practice or test run which required a certain accuracy level. For example, if the deck size was 200 simulated letters and accuracy required was 95%, no more than 10 errors could be made for the trainee to advance to the next speed or next lesson. The lesson was terminated automatically when the trainee got to 10 errors, the point where they could not possibly qualify on that deck. This was incorporated to avoid the situation where the trainee had to continue keying a deck even though the trainee knew that they could not qualify.

The error analysis feature of the CAI training consisted of an error analysis grid to be used by the student as well as the instructor. When a unsuccessful practice run occurred, the computer asked the trainee if an error analysis was desired. If requested, the errors were presented on the CRT in such a way that the letter preceding and following the error was also shown, as well as the number that should have been keyed and the number that was actually keyed.

Procedure. Individuals were notified by the Personnel Office to come to the office on a specified date if they were interested in employment as MPLSM operators. When the people reported, they were given an orientation to the Postal Service, administered an eye test, and scheduled for the first day of training.

The first group of trainees was randomly assigned to either the DETEX training or MICRO training prior to their arrival for the orientation. Since few MICRO trainees qualified on dexterity training in this group, subsequent call-ins assigned more trainees to MICRO than to DETEX.

The DETEX training is the conventional training used for MPLSM operators. The training device is a simulator of an actual MPLSM console, and DETEX cards simulate letters. Photoelectric cells on the training device read holes on the DETEX cards and deposit cards keyed correctly and incorrectly into separate bins.

The training program for MPLSM operators in the field test for both DETEX and MICRO training programs consisted of two parts, dexterity training which had an 18-hour maximum and outgoing primary training which had a 47-hour maximum. Dexterity training consisted of learning to key 3 digit numbers on the lower keyboard at 40 per minute. Outgoing primary training consisted of keying on the upper and lower keyboard at a rate of 60 letters per minute. Dexterity training had a maximum of 18 hours and outgoing primary had a 47-hour time limit.

The 47-hour maximum time limit became policy during the data collection phase of the study. The field test was exempted from this limitation and participants were allowed three weeks of training (approximately 22 hours) after the 47 hours if needed to qualify. The results of the field test are reported two ways, with and without the 47-hour limitation in outgoing primary training.

The tests used for qualification on dexterity training for both MICRO and DETEX training consisted of one run of 200 items of three-digit numbers at 40 per minute and 95% accuracy. The outgoing primary tests consisted of 250 items at 60 per minute and 98% accuracy.

On the day following successful completion on the outgoing primary qualification test, the trainee was given two days of MPLSM floor orientation, approximately two hours each day. On the second day of floor orientation, each trainee was given 20 minutes to key "live" mail on an MPLSM console. On the third workday after qualification, the individual was installed as an MPLSM operator and a job performance monitoring was begun.

Measurement of Job Performance . The North Suburban facility agreed to have each recently qualified MPLSM operator key a minimum of four hours each workday. The keying accuracy of each MPLSM operator was monitored by means of the EDIT procedure (Engineering Data Isolation Technique). For each MPLSM operator, six samples of 50 keyed letters each were taken, evenly distributed over each MPLSM assignment. The letters were then compared to a tape of the codes keyed for those letters to determine keying accuracy.

Data collection was performed by employees hired exclusively for this purpose. Daily assignments of the data collectors was controlled by a listing which identified each MPLSM operator's exact position for each 20 minute time period over the operator's tour of MPLSM assignment. Data collector activity was submitted to constant review by data collection assistants selected on the basis of their technical familiarity with MPLSM operations. Data collection assistants also reviewed tapes and samples for proper identification and scoring accuracy. Routine checks included random, unannounced reviews of samples scored.

Data was collected for a period of 12 weeks of job performance. Since the study had to be terminated before all trainees reached the 12th week, the samples for the later weeks drop off compared to the first weeks of job performance data.

Results

The overall pass rate for DETEX considering all qualifiers was 55.6 per cent (76.9% x 72.3%), while the overall pass rate for MICRO was 48.7 percent (51.6% x 94.3%). Considering only those who qualified on outgoing primary in 47 hours, the overall pass rate for DETEX was 26 per cent (76.9% x 33.8%), while the overall pass rate for MICRO was 40.9 percent (51.6% x 79.27%).

Results of the statistical tests indicated that DETEX took significantly less training time in dexterity ($t=8.22$, $p<.01$). In outgoing primary MICRO took less hours but the difference was significant only when considering all qualifiers ($t=4.64$, $p<.01$).

For total training time, the MICRO training took significantly fewer hours to train when considering all qualifiers ($t=2.30$, $p<.05$) but there was a nonsignificant difference when considering those who passed in less than 47 hours.

A comparison between DETEX and MICRO which contains data from pass rates and training time is the computation of the average number of hours used to successfully qualify one trainee. For dexterity, DETEX used 11.81 training hours to qualify one trainee while MICRO used 27.42 hours to qualify one trainee. In outgoing primary, the opposite was true. For all qualifiers, MICRO took 40.31 hours for one qualifier while DETEX took an average of 68.93 hours. Considering only those who qualified in less than 47 hours, DETEX took 116.96 hours while MICRO took 44.40 hours. The jump in hours for DETEX from 68.93 hours to 116.96 hours reflects the fact that 25 of the DETEX qualifiers (out of 47) needed more than 47 hours to qualify.

Comparisons for the MICRO and DETEX training groups for job performance error rates showed few significant differences over the 12 weeks of the collection of job performance data for the entire qualification group. DETEX trainees had the edge the first three weeks with only week 1 showing a significant difference ($t=2.96$, $p<.01$) while MICRO maintained an edge from weeks four through twelve. Only week 8 ($t=2.45$, $p<.05$) and week 9 ($t=2.59$, $p<.05$) showed significant differences. The same trend existed when the data analysis was restricted to those who passed in less than 47 hours.

Discussion

From a job performance standpoint, as measured by the first 60 days of job performance data, it could be concluded that neither training method exhibited any statistically significant advantage, although MICRO trainees generally performed better after the third week on the job.

The number of training hours needed to produce a qualified operator showed a distinct advantage of the MICRO training system. Given the existing condition of 47 hours maximum in outgoing primary training, MICRO needed 71.82 hours compared to 128.77 hours for DETEX. The reason for this difference was that MICRO failed many more trainees in the first part of training whereas DETEX failed individuals only when the 47 hours of the second part of training had been exhausted. Differences in the lesson plans between the two training systems may be largely responsible for this finding. In the dexterity training, MICRO used speeds up to 60 cards per minute while DETEX stayed at 40 cards per minute. Both training programs, however, used the identical test, administered at 40 cards per minute.

The conclusion from the pass rates, training time, and job performance data was that the MICRO/CAI training system was at least as effective if not more effective, than the conventional DETEX training. Additional advantages of the CAI approach was the ability to develop unique features in the training such as automatic speed increments, automatic termination of lessons after specified number of errors, increased feedback of error, and standardized work samples for qualification tests. The microcomputer-based system has the added advantage of each training system standing alone, thereby reducing the impact of hardware malfunctions. In addition, microcomputers do not require a controlled environment.

No doubt, the most distinct advantage of CAI training as applied to keyboard training is that of reduced cost. The conventional training system necessitates manual preparation of millions of DETEX cards annually because of wear and tear and changes in the addresses caused by changes in the carrier routes. In the CAI system, changes are accomplished by keying in changes on the disk, and this is needed only in the case of carrier route changes. Additionally, the cost of the microcomputer hardware is substantially less than the cost of the simulator currently being used by MPLSM training.

IMPROVED COURSE TEST DESIGN AND DEVELOPMENT

Edward N. Hobson, Ph.D.
DATA-DESIGN LABORATORIES
1755 S. Jefferson Davis Hwy.
Suite 307, Arlington, VA 22202

I. BACKGROUND

Training for the POLARIS Fleet Ballistic Missile (FBM) System was reviewed in the late 1960's and was found deficient in several respects. The actual training requirements were confusing, training pipelines were excessively long, overtraining was common, trained personnel often were unfamiliar with current technical publications, and no adequate method existed for evaluating training effectiveness. To correct these problems, the Chief of Naval Operations directed the establishment of the FBM Weapon System Training Program and assigned specific responsibilities for its implementation and administration. The advent of the POSEIDON and TRIDENT systems required further refinements to the training program and the name was changed to the Strategic Weapons System Training Program (SWSTP). The Strategic Systems Project Office (SSPO) was designated to implement and provide overall technical control of the program.

SSPO, in coordination with the Chief of Naval Education and Training (CNET) and his principal designee for submarine training, the Chief of Naval Technical Training (CNTT), has provided continuing support for development and implementation of this training program through a number of civilian contractors. Data-Design Laboratories (DDL) has been one of those providing assistance in curriculum materials coordination and management as well as in the evaluation component of the program. This paper describes the design and development of new end-of-course tests for use within the SWSTP. Much attention has been focused upon these tests because they are one of the most visible products of the evaluation component and they provide a great deal of the data for assessing how well individual courses and SWSTP, in general, are performing.

II. SWSTP OVERVIEW

The SWSTP is a systems approach to training composed of five major elements described below. For a more complete description of the training program refer to Proceedings 23rd Annual Conference of the Military Testing Association, Volume I, pages 191-201.

The Personnel Performance Profiles (PPP) are comprehensive, minimum requirements listings of the knowledges and skills required to operate and maintain a system, subsystem or equipment. The PPPs are essentially the result of hardware-oriented task analyses and are prepared using current information from approved engineering drawings, technical manuals, training literature, etc.

The Training Path System (TPS) assigns the knowledge and skill items of the PPPs to specific Navy personnel in a logical order and to a defined depth of knowledge and level of skill.

Curricula are composed of training materials designed to accurately reflect the training requirements identified in the TPS. Curricula may be designated as either formal or informal. Formal curricula are used in training facilities ashore to provide background, replacement, conversion, or advanced training.

Informal curricula are used in unstructured environments and frequently in on board training programs.

The Personnel Qualification Guides (PQG) are promulgated by the Submarine Force Commanders. The PQGs identify specific knowledge and skill requirements or standards that must be met by personnel to "qualify" or "requalify" for various watchstations on board a submarine or tender. These qualifications differ from training, in that they require a specific demonstration of ability using the actual equipment in its operational environment after appropriate training has been completed.

The Personnel and Training Evaluation Program (PTEP) is the element that measures, evaluates, and reports on the effectiveness of the total program. It is designed to assist management by monitoring, providing evaluation and feedback, and making recommendations for improvements. PTEP accomplishes its objectives by means of personnel testing, collection of test and nontest data, evaluation, and reporting. For a more detailed description of the PTEP element and how it functions refer to Proceedings 23rd Annual Conference of the Military Testing Association, Volume I, pages 573-579. Most of the day-to-day responsibility for implementing and operating PTEP rests with the Central Test Site (CTS) located at Dam Neck, Virginia.

III. NEW END-OF-COURSE TEST DESIGN AND DEVELOPMENT REQUIREMENT

The SWSTP utilizes approximately 180 courses spread over five enlisted communities. These courses of instruction range from one week to twenty-six weeks in duration. A typical technician might attend up to twelve of these courses. End-of-course tests originally prepared for these courses were adequate, but areas for improvement were evident. Consequently, an effort was initiated to develop a set of test construction procedures that would result in better tests--tests which focused most of the questions upon aspects of the training course deemed to be most important by subject matter experts.

After the new procedures had been developed, a limited test (validation) was initiated to ascertain: 1) how well the new procedures worked, 2) how much manpower/resources they consumed, and 3) if there was an improvement in the data received from the new tests.

IV. STATUS OF END-OF-COURSE PTEP TESTS

During the years 1980 and 1981, efforts were made to improve the quality of tests provided by CTS to the SWS training activities for administration as comprehensive end-of-course tests in advanced and replacement training courses. Under the direction of the SSPO, DDL designed procedures for the development of new end-of-course tests which would provide a better assessment of student mastery of the more important learning objectives.

The primary purpose in designing and producing improved tests was to provide better assessments of how well students had mastered the most important aspects of the course and how well the curriculum met the training requirement. When PPP tables were developed the system lacked some of the rigor now deemed appropriate. Therefore, the tests designed using the equipment orientation of the PPPs without sufficient attention to "the most important objectives", gave emphasis to hardware knowledge instead of functional operations and maintenance. Thus, end-of-course tests based upon these sources (i.e., the PPPs, the

amount of time devoted to topics in the classroom, and at times the availability of good test items) resulted in reasonably good tests but with an emphasis on knowledge of hardware rather than functional operations and maintenance tasks.

This new look at the test design and development process refocused attention upon the fact that evaluation involves the whole system. One of the first shortcomings discovered was in the objectives. Frequently there was no hierarchy of objectives within major tasks. For example, the four courses selected for the validation effort are listed below with the number of Section Learning Objectives (SLO) contained in each. The SLO was chosen by test design personnel as the most appropriate curriculum level for testing.

Course	Number of SLOs
D-to-D CONVERTER	22
MTRE MK 6 MOD 3	17
NAVAIDS BLOCK 3	40
LAUNCHER BLOCK	200

In the test case when the number of SLOs exceeded 40, subject matter experts were unable to reach agreement on which SLOs were most important. For the validation effort the Launcher Block test was split into two parts. This permitted a larger number of SLOs to be tested and served as an interim solution. Subsequently, courses that seemed to have an excessive number of SLOs were reviewed by representatives from the training facilities, CTS, and DDL. These working groups restructured, reworded, and revised the objectives so that the intent of all objectives was retained while building a hierarchy to facilitate both instruction and testing. Furthermore, most of the objectives lacked a criterion of satisfactory performance. While these facts were known, in general, the impact was not really felt until more rigorous evaluation (i.e., better tests) procedures were developed and tested.

V. COMPARISON OF NEW AND OLD PROCEDURES

Tests designed using "old" procedures focused upon an analysis of applicable curriculum and OAC, on evaluation of associated knowledge and skills contained in the Instructor Guide, and on the length of a course of instruction. Usually tests were designed to meet the following statistical and time requirements:

- o Not more than one test item per hour of instruction
- o Short courses contain a minimum of 30 test items
- o Knowledge test areas contain a minimum of 5 test items
- o Total average work time for all skill exercises administered in one three-hour session should not exceed two hours
- o Total test time (all knowledge sections) will not exceed three hours per course
- o Combined knowledge and skill test time will not exceed six hours

The old end-of-course development process was initiated 28 weeks before the test was due to be administered. New end-of-course tests require about the same amount of time for design, development and production. However, there are some significant differences between old and new procedures that provide a better test of the "most important objectives" of the courses. These differences are presented in Figure 1.

PROCESS	NEW	OLD
TEST DESIGN	3 SMEs are required. Design is based on review of actual subject matter as stated in curriculum SLOs. SLOs are prioritized to ensure testing of most critical learning objectives. Process reveals discrepancies between the IG and what is taught.	1 SME is required. Design is based on OACs, specified only by PPP item & TOS level. Except for identifying instruction times & sequence, no specific requirement exists to review IG.
TI SELECTION AND REVIEW	2 SMEs are required. TI content is matched with SLO content (to level of TLO & DP, if needed). A rigorous review (technical applicability, format, training specification coding, grammar) is required for those TIs which match the SLOs. Detailed procedures flag faulty TIs for review/improvement.	1 SME is required. TI selection is based on PPP/TOS test design criteria. Procedures do not specifically require verification of PPP/TOS applicability previously assigned to selected TIs. Procedures for review of TI are not detailed.

ACRONYMS

DP - Discussion Point
IG - Instructor Guide
OAC - Item-to-Topic Objective Assignment Chart
PPP - Personnel Performance Profiles
SME - Subject Matter Expert
SLO - Section Learning Objective
TI - Test Item
TLO - Topic Learning Objective
TOS - Training Objective Statement

FIGURE 1. Comparison of New and Old End-of-Course Test Procedures

VI. ADMINISTRATION OF NEW TESTS

Upon completion of the design, development, and review of the tests, they were administered to students by Navy instructors without mentioning that they were taking a different type of test. Following the test, each student was asked to complete a short questionnaire. The numbers of students ranged between 30 and 52 for each of the courses in the validation effort.

Student test results and information from the questionnaires were used in subsequent analyses. Navy instructors also completed a questionnaire following the last administration of the test.

VIII. DISCUSSION OF FINDINGS

The new procedures improve the design and development process so that the school is more certain that those students who score high on the new end-of-course tests have attained reasonable mastery of the most important objectives. Notice that an unequivocal statement of mastery is not made. Why? Because many of the objectives (some developed several years ago) do not have specified criteria or conditions under which performance will be measured. The new procedures do, however, provide an improved test. Furthermore, these procedures can be used for the design and development of criterion referenced tests in the future.

In the process of developing and validating the new procedures, a number of items surfaced. For example, in the four courses validated some of the "actual instruction" in these courses was found to differ from the approved curriculum. It became clear from this information that the SWSTP revision/update process had failed in these cases. The curriculum control group is aware of this fact and is making a more careful surveillance of the curriculum.

During the validation of the end-of-course test procedures, it was also determined that the test item bank did not support certain objectives. These shortages tended to be associated with test items for "operations" and "maintenance" areas. Most "knowledge objectives" were adequately supported and occasionally there was an overabundance of these test items. This discovery led to the development of new test items to support the four validation tests.

Now, as new "end-of-course" tests are designed by CTS using the new procedures, the total shortfall in test items is being determined. Additionally, a new test item acquisition form has been developed. This new form provides detailed information on objectives for use by the test item developer. Each new acquisition form also has a copy (listing) of current applicable test items attached so that new test items will not duplicate what is already in the test item bank. Test items acquired in this manner will match more closely the Navy's testing needs.

It had been known prior to the validation effort that certain deficiencies existed within the 22-25,000 bank of test items. Some of these problems were with the primary technical data of individual test items or with the related data--that data/document referenced by the test item. Other deficiencies known to exist included test item stem faults, distractor (plausible incorrect answers) shortcomings, and grammatical errors. However, until these new procedures were developed, the resolution was basically unmanageable or too costly. Therefore, while new end-of-courses tests are being developed using

the new procedures, a very serious problem (test item bank deficiencies) is also being resolved.

The validation effort also pointed out that the Navy should require all future courses (new and revised) to have a hierarchy of objectives. This requirement alone will facilitate future evaluation efforts since tests can be designed to sample the higher level objectives rather than having to sample from an overabundance of objectives at a lower level of difficulty.

Questions regarding acceptable performance criteria have resurfaced and are being reviewed by CTS, SSPO and CNTT. Likewise, the scoring, reporting, interpreting, and presenting of test results to the submarine crews are also being re-examined.

IX. CONCLUSION

The preceding discussion of the findings of the validation effort should not leave one with the impression that previous evaluation/testing efforts were inadequate. In fact PTEP has served the SWSTP well as the evaluation component. Prior end-of-course tests satisfied the basic testing requirements. However, the time had come to upgrade and improve the testing aspect of PTEP's functional responsibility. The problem was "how to" upgrade an operational evaluation system with no increase in funding and with as few disruptions to the system as possible. It was not an easy undertaking. The newly developed "end-of-course" procedures turned out (hindsight) to be an excellent vehicle to initiate change (improvements) that attacked basic problems while generating only a moderate (manageable) set of attendant problems.

In summary, the new end-of-course test design and development procedures take no more total time than the previous test development procedures and provide a test much more closely related to training objectives than the old tests. At the same time they have opened up other areas for improvement; curricula development--hierarchy of objectives, better test item procurement process, and potentially better scoring and reporting of the data. While the procedures worked well for courses with a small (40 or less) number of objectives, the results were not as conclusive for courses with a large number of objectives all of which were at a comparable level of importance.

The validation effort also left some unanswered questions. For example, should data from tests using a criteria base line receive the same analytical treatment as the scores from tests designed to discriminate or, when applied to a different test instrument, produce a bell-shaped curve? On one hand, having all students make correct responses to all questions would be considered excellent--everyone would have met the criteria. On the other hand, these same test results would create a problem where the object is to provide some discrimination between individuals taking the test--no bell-shaped curve. Furthermore, once the question has been answered regarding the appropriate analytical treatment the implications for computer software must be considered. Likewise, any scoring trend analyses must recognize the potential step change that could be caused by moving toward a more criterion-referenced testing system. Still another question deals with the whole scoring scheme (pass/fail, met criteria/did not meet criteria, scores by objectives, percentiles, T-scores, etc.). So, while use of the new procedures is producing better end-of-course tests, a number of important questions and issues still need to be addressed, considered, studied and resolved.

MISSION BASED AND ISD BASED COMPONENTS OF TASK CRITICALITY JUDGEMENTS

R. Gene Hoffman and Eugene H. Drucker
Human Resources Research Organization
and
David W. Bessemer
Army Research Institute for the Behavioral and Social Sciences

Given the limited time and resources available for soldier training, there is always the need to be careful in the selection of tasks to be included in formal institutional training. The general concept of task criticality arose to guide that selection process. Tasks which are more critical receive training priority. Less critical tasks are included in training only if time and resources allow.

Determination of task criticality is problematic for two major reasons. A frequently considered issue concerns the dimensions on which task criticality is judged. Another and related issue concerns the context within which task criticality is considered. Both issues will be addressed in this paper.

Before continuing, a significant assumption must be made clear. Assessment of task criticality is almost of necessity a matter of judgement. Empirical determination would require comparison of mission success with and without performance of each component task, or with the quality of performance of each task varied over a number of attempts at the mission. Obviously, the feasiblity of such manipulations is inconceivable and, particularly for combat missions, the ethics reprehensible. Therefore, expert judgements usually comprise the data base for determining task criticality. Certainly one can ask about the validity of these judgements, but it is a question that cannot be directly answered in most cases. Criticality judgements may be accepted as judgements and often the best available data. Our best safeguard is to obtain average rating from as large a sample of expert judges as feasible in order to control for individual idiosyncrasies. On the other hand, the effects of the methods and procedures used for obtaining these judgements is an open question.

TRADOC Circular 351-4, Job and Tasks Analysis (1978) defines a critical task as "as task which is essential for: Accomplishment of the unit's mission or successful individual job performance or survivability in combat situation" (p.H-4). Consequently, Circular 351-4 also suggests that ratings be gathered concerning the consequences of inadequate performance of potential training tasks. In addition, Circular 351-4 prescribes the use of three additional judgements that are particularly germane to the training perspective. One judgement concerns the time available to begin a task once the need for it arises. Implicit in this consideration is the idea the delays may be used as preparation time to acquire or develop the needed skill. Learning difficulty is another criticality factor prescribed by the ISD model. The final factor is percent performing a task. The greater the number of soldiers required to do a task, the greater the benefit of standardized institutional training.

Circular 351-4 also specifies that a weighting system be established and used by each proponent school to combine these criticality factors into an overall judgement of each tasks' criticality. It also implies that such weighting systems should not be rigidly adhered to. Thus, to some extent judges' overall assessments of task's criticality for mission success may play a significant role in determining what tasks are eventually included in training. One purpose of this paper is to examine the relationship of global ratings of task criticality for mission success to three of the ISD prescribed judgements. These include time available to start the task, time to learn, and the amount of damage or injury possible from failure to perform the task. That is, how are experts' judgements of ISD model components related to their assessment of overall task criticality?

The accomplishment of any mission depends on numerous interdependent functions, activities, and events. At one level of abstraction, it is very easy to argue that some tasks must be more important than others. On the other hand, it is also easy to construct situations where even the most minuscule tasks not performed can lead to mission failure (e.g., the proverbial nail for the horse's shoe). The possibility of these situational differences are often recognized by training developers and expert judges. However, because any given training program is designed to prepare personnel to perform in a variety of situations, judgements concerning task criticality are required without reference to any special circumstances. Rather, they are made in the abstract.

There is no reason criticality judgements cannot be tied to specific situations. Another part of the research (Drucker, Hoffman and Bessemer, 1982), compares differences between summary ratings of task criticality made with and without reference to specific mission scenarios.

When considering the criticality of a task within the context of a combat scenario, the criticality dimensions germane to training (i.e. time to learn and time available) seem less important. Rather, the relationship between the tasks and combat functions seems more in line with the basic definition of a critical task. That is, tasks can be rated for their criticality in relation to unit (1) fire power, (2) mobility, (3) command and control, (4) sustainment of effectiveness, and (5) survival of men and equipment, as well as their overall contribution to the success to specifically described scenarios. The second purpose of this paper is to look at the relationship among expert ratings of these factors.

To acquire combat criticality ratings, four combat scenarios were constructed. Scenarios included action on contact, hasty attack, occupy battle position and defend battle position. Because of differences in the objectives of these missions, the combat functions may have differential importance. To investigate this, relationships among criticality for these functions and overall mission success will be examined separately for each scenario.

The final question of this study concerns the comparison of ISD based ratings to combat mission based ratings. That is, are judges' ISD component ratings representative of the mission based global criticality ratings? Similarly are judges' combat function ratings related to ISD global criticality ratings?

Method

Questionnaire

Two separate questionnaires were developed. One questionnaire was based on ISD components and used to assess 161 tank platoon leader tasks. Ratings were obtained for TIME TO LEARN (none, one hour or less, several hours, one day, two or more days), TIME AVAILABLE (none, one minute or less, several minutes, several hours, one day or more), amount of DAMAGE or injury (none, small, moderate, large, extreme), and the overall effect of task performance on successful accomplishment of the team mission (none, small, moderate, large, extreme).

The second questionnaire assessed the criticality of a subset of these tank platoon leader tasks relative to combat functions within the context of the four combat scenarios. The scenarios were further divided into mission phases. That is, hasty attack consisted of two phases, conduct fire and maneuver, and conduct the assault. Eight tasks were included in conduct fire and seven tasks in conduct the assault. The action on contact scenario included three phases: (1) immediate action (three tasks), (2) develop the situation (three tasks), and (3) occupy suppressive fire position (six tasks). Occupy battle position consisted of occupy platoon battle position (six tasks), and organize platoon battle position (17 tasks). Defend battle position included three phases: (1) maintain surveillance in platoon sector (two tasks), (2) initiate indirect fire in platoon sector (two tasks) and initiate direct fire in platoon sector (nine tasks). Fifty-one different platoon leader tasks were included in the combat mission based questionnaire. Six tasks were repeated in more than one scenario or in more than one phase within the same scenario for a total of 66 unique tasks within scenario phase combinations. All of these combinations were rated for the tasks' effects (none, small, moderate, large, extreme) on FIRE POWER, MOBILITY, COMMAND, SUSTAIN, SURVIVE and OVERALL SUCCESS.

Subjects and Procedure

Two groups of US Army Armor Officers enrolled in the US Army Armor School's Armor Office Advanced Course were administered the questionnaires. One group (n=65) completed the ISD based questionnaire. The other group (n=57) completed the combat mission based questionnaire.

Criticality ratings for each task on the four ISD scales and the six mission based scales were averaged across raters. These mean ratings constituted the data base for this paper. Interrater reliabilities were calculated using Cronbach's alpha with tasks treated as subjects and raters treated as items. Reliability estimates ranged from a low of .86 for sustainment ratings from the mission based questionnaire to .96 for time available and for time to learn for the ISD based questionnaire.

Results and Discussion

The ISD model suggests that time to learn, time available to start the task and the anticipated amount of damage or injury from failure to do the

task are key components of a task overall criticality. On the other hand the scenario model suggests that a task's effect on unit fire power, mobility, command and control, survivability and sustainment are key components of task overall criticality. Multiple regression was used to examine the relationship among these criticality ratings. ISD correlations among components were based on the ratings of 161 tasks. Because of the repetition of task within the scenario model and the expectation that task ratings might differ by scenario and scenario phase, each task-within-scenario phase combination was treated separately and the correlation based on an N of 66 unique task/scenario phase combinations.

For the 161 platoon leader tasks, the ISD component DAMAGE is highly correlated with success ($r=.84$, $p<.01$) TIME TO LEARN is also strongly correlated with success ($r=.61$, $p<.01$); tasks which take longer to learn are most likely to have a significant effect on mission success). However, TIME AVAILABLE shows essentially no relationship to SUCCESS ($r=.08$, n.s.). The regression analysis, Table 1, collaborates these conclusions. In the full model with all three components, TIME AVAILABLE did not contribute significantly to the regression equation. DAMAGE and TIME TO LEARN were retained with statistically significant betas in a reduced model. The multiple $R^2=.74$ between these two scales and SUCCESS is greater ($p<.01$) than the zero order of either DAMAGE or TIME TO LEARN. However, the unique contribution of TIME TO LEARN, represented by an increase in R^2 from .70 to .74, is certainly not large. The regression weights also indicate that DAMAGE contributes substantially more to ratings of a task's effect on SUCCESS than does TIME TO LEARN.

Table 1
Regression Analysis of Successful
Accomplishment Ratings

ISD Components:	Standardized Regression Weights: Damage	Time to Learn	Time Available			R	R^2
Full Model:	.74**	.19**	.06			.86**	.75**
Reduced Model:	.71**	.24**				.86**	.74**
Scenario Components	Fire Power	Mobility	Command	Survive	Sustain		
Full Model	.44**	.14*	.14*	.08	.43**	.93**	.86**
Reduced Model	.43**	.15**	.11*	-	.50**	.93**	.85**

*$p<.05$
**$p<.01$

The regression analysis for the scenarios model scales is also presented in Table 1. SURVIVE was substantially correlated with SUSTAIN ($r=.90$, $p<.01$) and in the full model with all fire predictors included, only SURVIVE failed to make a statistically significant contribution. The four remaining components achieved a remarkable .86 multiple R^2 with successful accomplishment with fire power and sustainment displaying the largest

regression weights. Clearly, either set of component ratings, ISD or combat based, provide adequate representation of judges' global ratings of task criticality. However, the differences among the component weights suggests that they are not equal in importance. Judgements about tasks' criticality related to platoon fire power, sustainment, and minimizing damage and injury appear to predominate judgements about a tasks overall criticality.

To examine the possibility that the contributions of the scenario components may vary across scenarios, a regression analysis was conducted to examine the interaction between scenario and components. The set of scenario by task interaction terms that were entered into the regression solution significantly ($p<.01$) increased the R^2 from .92. To uncover the specific nature of the interactions, separate regressions were calculated for each of the four scenarios. Reduced equations for these separate regressions are presented in Table 2.

Table 2
Regression Analyses on Successful
Accomplish for Each Scenario

Scenario	Standardized Regression Weights				R
	Fire Power	Mobility	Command	Sustain	
Action on Contact	.93**	-	-	-	.93**
Hasty Attack	.93**	-	-	-	.93**
Defend Battle Position	-	.35*	-	.66**	.99**
Occupy Battle Position	.66**	.28*	-	.40*	.93**

*p < .05
**p < .01

The regression weights do appear to vary substantially across the four scenarios. Fire power is the only combat function which enters the two essentially offensive scenarios (Action on Contact and Hasty Attack). On the other hand, mission success for Defend Battle Position appears to be primarily dependent on sustainment of effectiveness in the judgement of Armor officers. The Occupy Battle Position scenario, which has elements of offense and defense, includes both fire power and sustainment as important task attributes.

The theoretical significance of the variation in these weights is diminished when correlations between the overall reduced composite and SUCCESS ratings are examined within each scenario. That is, when the predictions of SUCCESS ratings are made using the regression equation derived across all four scenarios, and these predictions correlated with SUCCESS within each scenario, the correlations are .90, .94, .97 and .90 for Action on Contact, Hasty Attack, Defend Battle Position and Occupy Battle Position, respectively. There correlations certainly approximate the multiple R's for the regression equation developed within each scenario. This suggests that the relative weights among the composites are less important than the components themselves. That is, the multicollinearity among the components is large enough that shifting the relative size of the regression weights does little to affect the overall predictability of SUCCESS judgment by the set of components.

Cross-Method Comparisons

The final correlational analysis involved a validation of both the ISD and Scenario rating component regression composites against the SUCCESS ratings made by the alternative method. Each of the 66 task-within-scenario-phase combinations from the scenario method, were matched with the ratings from the ISD method. Repeated tasks simply received the same ISD values while their scenario ratings could vary across the scenario in which they appeared. Composite indices were calculated from the reduced regression equations previously developed for each method. Correlations between the composite indices and SUCCESS ratings were calculated for the set of 66 task-within-scenario combinations.

The ISD composite correlated with scenario SUCCESS almost as highly as with ISD SUCCESS (r=.69 compared to r=.76) for the subset of tasks included in this analysis. The combat based composite showed a more noticeable decrease (r=.45 with ISD SUCCESS compared to r=.93 with combat scenario SUCCESS). The correlation between the two SUCCESS ratings was .57 (p<.01), and the correlation between the two composites was .67 (p<.01).

This pattern of results suggests that the combat based criticality ratings of tasks' contribution to mission success relate more highly to combat functions than do the ISD criticality ratings. A supplemental analysis revealed that when combat SUCCESS was regressed on the ISD components (using the tasks common to the two methods), DAMAGE was again the dominant variable with TIME AVAILABLE instead of TIME TO LEARN as the second and only other variable (total R=.71). Given that TIME TO LEARN is a training factor, while TIME AVAILABLE is a combat situation factor, this is further support for the argument. Furthermore, regression of ISD SUCCESS onto the combat scenario components did not yield a composite with any more predictive power than the original combat composite (R=.47, with FIRE POWER as the only variable entering, p<.01). This argument should not be construed to mean that the ISD ratings are invalid. Rather, we know less well what they mean. Perhaps our efforts would be furthered if a more clear separation were made between task criticality as it relates to mission success, and training criticality, which relates to training management factors. Clearly, combat scenario based ratings can increase our confidence in task criticality ratings for mission success.

References

Drucker, E. H., Hoffman, R. G., & Bessemer, D. W. A comparison of two methods for assessing task criticality. Military Testing Association Proceedings, 1982.

TRADOC Circular 351-4, Job and Task Analysis, 1978.

Does Measurement Hurt or Help a Sacrosanct System?

Clark L. Hosmer
39 Longwood Drive
Shalimar, Fl 32579

~~My purpose~~ The purpose of this paper is to examine whether the Honor System at West Point would be hurt or helped by making the cadets aware of survey findings of how they stand on toleration of dishonesty.

The Honor Code states, "A cadet will not lie, cheat, or steal nor tolerate those who do." (USMA p. 1). If a cadet is found guilty of dishonesty, he or she is dismissed. Similarly, if a cadet is found guilty of tolerating another cadet's dishonesty, the tolerator is dismissed.

The Honor System is Sacred

The renown and sanctity of the Honor System depends upon its rigor. Dismissal of cadets who are found guilty of personal dishonesty, or toleration of another's dishonesty, resembles excommunication by a church of its unworthies who are cast out from the in-group of communicants. Remaining members of the in-group are thus confirmed as worthy of continued service. In sum, honor is revered as an all-or-none proposition. The tenet of all-or-none is shared by the United States Air Force Academy's Honor System (USAFA p. 11).

Only about one percent of the cadets may be found guilty and dismissed in a year. As to toleration, over a period of ten years, fewer than one in one thousand cadets was found guilty of toleration alone. (Borman p. 17). The Honor System apparently has accomplished the awesome job of convincing young men and women that non-toleration of others' dishonesty must be put before loyalty to closest friends.

Before West Point, youngsters grow to know that the worst crime in the book is to "rat on a buddy." At West Point, that peer loyalty is further reinforced by close support of classmates in joint tasks. New cadets are taught, however, that the higher loyalty is their responsibility to the Honor Code and its non-toleration clause. The central obligation in the motto of West Point - Duty Honor Country - and each cadet's oath that service to the nation is more important than self or friends is fulfilled by the vast majority of the cadets. In sum, the Honor Code is held to be an absolute and is revered as sacred.

But Honesty is a Variable, Not an Absolute

Dishonesty may range from signing a false official certificate to the white lie of a cadet flattering his girlfriend, from using notes taken into an examination to using information accidentally heard in a social conversation with a friend who has taken the exam before, from stealing a stereo set to

using a ball point pen in a government office, walking away with it, and later keeping it as an item not worth making a special trip to return it.

Similarly, a cadet's information about another cadet's dishonesty is a variable, not an absolute. Information may range from direct observation, to hearing another cadet's shocked statement that he was flabbergasted when he saw his good friend cheat on an exam, to simply hearing about somebody in the next company having stolen money. Presumably, personal observation is ground for reporting an honor violation. But what to do about circumstantial evidence, or compelling hearsay as cited for example, or persistent rumors of group cheating?

Finally, a cadet's idea of a reportable violation is a variable, not an absolute and thereby toleration may vary. Toleration may range from being an accessory before-the-fact, to counseling the violator but not reporting the violation and thereby becoming an accessory after-the-fact, to tolerating a good friend's theft of a government ball point pen because dismissal is thought to be too gross for such an offense.

Problems in the Honor System

In the last 31 years, cheating scandals have occurred six times. More than 100 cadets were dismissed in the latest episode. On earlier occasions of group cheating, 19 to 90 cadets were dismissed. Some people say that proves the system works. If a group is found guilty, out everybody goes. While the rigor of handling those brought to dock is impressive, what was the basic cause of the half-dozen large-scale cheating scandals?

During the investigation of the latest episode, both cadets and officers cited views that only a fraction of the cheating that was 'known' was reported. (Borman p. 15-17). An official survey revealed that more than two-thirds of the cadets said that they would not report a good friend for a possible honor violation and more than one-third said they would not report a good friend for a clear-cut violation. (Borman p. 14).

The Government Accounting Office (GAO), citing the Superintendent's Honor Review Committee study that was completed before the latest group cheating was discovered, reported that the non-toleration clause was, "one of the biggest problems for the cadets." The GAO also reported that, "Some cadets feel that friendship is more important than reporting a fellow cadet," and, "Generally, toleration increases as a cadet progresses through his four years." (GAO p. 56).

Finally, because toleration is held to be as serious as personal dishonesty, investigation of an honor violation naturally should look into whether other cadets tolerated the offense. Therefore, the almost total absence of convictions for toleration seems strange.

In sum, the heart of the vulnerability of the Honor System to group cheating may be cadet toleration of the few individual honor violations that occur.

Indeed, the evidence suggests that toleration of toleration was widespread. So far as group cheating is concerned, its growth is associated with the pressure on a cadet to join a violator he has tolerated. Both may be dismissed. There's little difference between being hung as a goose or as a gander.

Three other problems merit brief consideration. The Honor System can be used to enforce regulations. A cadet may be put upon his honor to say if he has shaved instead of the inspector deciding whether the cadet is acceptably whiskerless. Cadets tend to resent many requirements being checked for under the Honor System as an inspector's quick and easy way to insure compliance. The result can be reactions of technical compliance but with clever ways to beat the system.

Another problem is a history of cases of heavy handling of what can be considered trivial or remediable offenses. For example, a cadet wore the coat of an upperclassman to a movie that he was not authorized to attend. He may have accepted the risk of breaking a regulation but was dismissed as a dishonorable person. Another cadet was dismissed after he reported himself for having said he had shaved but he had not shaved. (Borman p.6). Similarly, a cadet was dismissed after reporting himself for stating that he had done ten pull-ups but he done only two. (Borman p. 21).

A fourth and last problem is whether toleration is a matter of personal honor as its inclusion in the Honor Code implies. Or, is non-toleration strictly "an awesome duty" as the official text on the Honor System states in the section on the philosophy of non-toleration? (USMA p. 15). Surprisingly, the official survey of the Corps of Cadets showed that 45% said they wanted toleration removed as an honor violation. (Borman p. 14). Perhaps the Corps' exploration of all of the pros and cons of defining toleration as "dereliction of duty" without any change in the Honor Code would be enlightening.

Proposed Use of Survey Findings

In the fall of 1981, I submitted a proposal to the superintendent of West Point, "To increase the effectiveness of the non-toleration policy." If acted on, the proposal could have produced something along the lines of Figure 1. The graph shows the percent of cadets, by class, who are willing to report a good friend for an honor violation. The questions proposed were the same as used in an earlier official survey, "Would you report a good friend for a clear-cut honor violation?" and "Would you report a good friend for a possible honor violation?"

The hypothetical results are consistent with the GAO report that a cadet's inclination to tolerate another cadet's honor violation increases as the cadet progresses through the four years. (See Note on References page)

H Y P O T H E T I C A L views of non-toleration

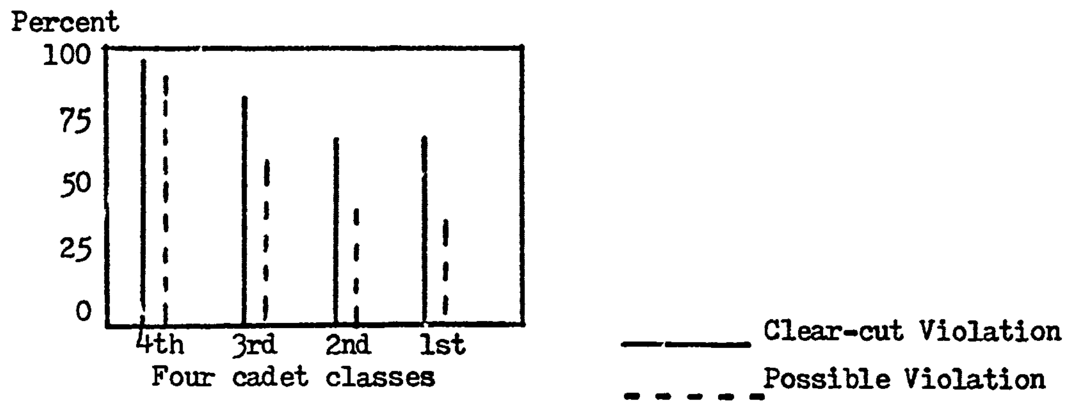


Figure 1. Percent of cadets willing to report[1] a violation[2] by class.

1. "Report" means that, to enhance validity, the alleged violation is checked with the violator, then reported to the Company Honor Representative.

2. An alleged "violation" may be observed or suspected.

H Y P O T H E T I C A L views of non-toleration

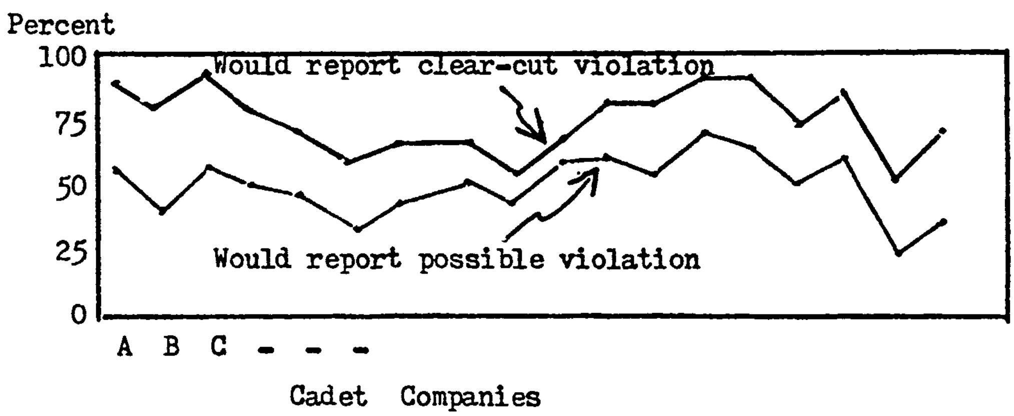


Figure 2. Company differences in percent of cadets willing to report an alleged honor violation.

Company variation may highlight potential trouble. Some companies may need free-for-all airing of uestions, preferably in small groups of peers. Elney found that uninhibited discussion among peers to be an effective way for emergence of agreed loyalty to group goals as contrasted with individual competitive interests. (Elney p. 84). Role-playing can be effective in helping cadets to learn how to confront a friend with tact and persistence to validate a suspected violation. That skill requires maturity and can be developed but not by lecture and exhortation. In sum, an overall average plus differences among organizations may point to problems and serve as yardsticks to reflect the effectiveness of remedial actions.

To help see whether the other three problems exist, my proposal included: (1) exit interviews of every cadet dismissed for dishonor, academic deficiency, or other reason, (2) A paper from every cadet once each year on any self-selected strength or limitation of the Honor System, and (3) views of exchange students from the Air Force and Naval Academies because they have familiarity with two honor systems.

Would Cadet Knowledge of Research Findings Hurt the Honor System?

A friendly critic of my proposal said that if cadets were to know that the levels of toleration were high, West Point would be taking an enormous risk. The non-tolerating cadets might join the tolerators! Moreover, would not leaks of the findings to the public media produce a scandal of itself? The old grads would see solid evidence that the Corps has gone to hell.

Would Cadet Knowledge of Research Findings Help the Honor System?

If alarming levels of toleration were revealed, what better foundation is there than objective estimates of the problem areas? With all their slippages, well conducted surveys can provide estimates superior to subjective impressions of the workings of the system. As to leakage of findings to the public media, the record of the furor over the latest cheating scandal included staunch defenses of West Point's splendid reputation for the Honor System and editorial confidence that the reputation soon again would be earned, as it has been.

Finally, who owns the Honor System? The cadets do. The officers in the academy and the old grads think that they own a part of it because the Honor System has had such a profound influence in their lives. Nevertheless, the cadets know that the Honor System is theirs to nurture and to hold the new cadets to understand, comply with, and revere. On that ground, I think that the odds favor good things happening if the Corps of Cadets were provided findings from research on the workings of the Honor System. As a former superintendent said, in the long run, openness as well as honesty is the best policy.

In conclusion: (1) Available data support the idea that toleration of honor violations is associated with group cheating. (2) Organizational and class differences in willingness not to tolerate honor violations help to identify problems in the system. (3) Honor Committee instructional focus on the complexities of implementing the non-toleration policy, small discussion groups among peers to air questions, and role-playing to develop skills in confronting a suspected violator would help to solve the problems of individual cadet implementation of the non-toleration policy. (4) Exit interviews of dismissed cadets, annual papers from every cadet on self-selected aspects of the Honor System, and views of exchange students would help to describe other possible limitations of the Honor System.

More power to the sacrosanct Honor System at West Point!

REFERENCES

Borman, F., Johnson, H.K., Pye, A.K., Tate, W.M., Walker, J.Y., and Wilcox, H.S., Report to the Secretary of the Army, Assembly, March 1977.

Edney, J., Free Riders En Route to Disaster, Psychology Today, August 1979.

United States Air Force Academy, Honor Code Reference Handbook of the Air Force Cadet Wing, June 1982.

United States General Accounting Office, Academic and Military Programs of the Five Service Academies, 31 October 1975.

United States Military Academy, The Honor Code and the Honor System, USCC Pamphlet 632-1, June 1982.

Note: With regard to tolerance increasing during a cadet's four years, after a new cadet has carried the non-toleration torch for a year, perhaps honesty begins to loom as a variable. The harder a cadet finds honesty to be an all-or-none proposition, the more readily he or she, on graduation, may phase into the responsibilities of an officer. Officers operate with less than puritanical correction of others' every lapse from rectitude. Moreover, while serving with integrity, an officer often works in a sea of classified information. The truth is told to those who have an official 'need to know' the truth. In intelligence work, quiet forms of deception are often part of the job. All this does not mean that upperclass cadets may be excused for tolerance nor do cadet and officer need have less regard for the power and the beneficial influence of the Honor System.

For their criticisms of this paper, I thank John Carley, Patricia Dawes, Charles Hosmer, William Jones, Dan and Fran McElheny, Richard Morgan, M. A. Roth, Brett Sortor, and Teresa Ward.

AD P000881

IDENTIFICATION OF RATING POLICIES IN TRAINING EMPHASIS TASK FACTOR DATA

HANS P. JANSEN, SQNLDR, RAAF

AIR FORCE HUMAN RESOURCES LABORATORY (AFHRL)
MANPOWER AND PERSONNEL DIVISION
BROOKS AFB, TEXAS 78235

INTRODUCTION

The Air Force Occupational Measurement Center (USAFOMC) conducts task based occupational surveys of Air Force specialties which include the collection of specialty supervisors' ratings on task factors such as training emphasis (TE). These TE ratings are utilized to determine the first-term task training priorities for individual specialties. For a small number of specialties TE ratings have been difficult to interpret due to poor rater agreement. Current analysis technology does not permit these complex data to be fully processed and applied. The suggestion has been that the data for these "complex specialties" may contain information limited to job areas within a specialty needing special training attention. This paper reviews the currently employed analysis technique, discusses possible causes of poor rater agreement and reports research findings for the effect of sample size on rater agreement and the utility of employing cluster and factor analysis techniques for identifying multiple rating policies in training emphasis data.

BACKGROUND

Analysis of training emphasis ratings is usually performed using REXALL, a special purpose program within the Comprehensive Occupational Data Analysis Programs (CODAP) system. The two main functions of REXALL are: (a) to calculate the mean training emphasis for each task, and (b) to assess the level of rater agreement. With respect to rater agreement, REXALL is designed to cope with a sample of raters who are anticipated to be relatively homogeneous in terms of their rating ability and judgements.

Ratings for TE (first-term training emphasis recommended) are made against a nine-point scale; 1, extremely low to 9, extremely high. However, the instruction to "rate only tasks which you believe require training for first-termers" recognizes the validity of a zero rating. By default, all non-ratings are interpreted as zero ratings equating to "no training recommended" and are included in calculating the mean TE for each task. Two further consequences of the zero rating are that: (a) the zero anchor point is perceived as distorting the meaning of the 1- to 9-point relative ratings, and (b) the dichotomous recommend training/recommend no training decision skews the distribution of task means towards zero. These two factors prevent standardization of the ratings as a means to reduce rater differences.

As a measure of rater agreement, REXALL computes two indices of interrater reliability: R_{11}, single rater reliability which approximates the average of all possible pairwise rater correlations; and R_{kk}, reliability for a sample of k raters, which is the expected correlation between the set of observed sample task means and the task means of an hypothetical equivalent sample. R_{11} and R_{kk} meeting or exceeding minimum criterion values are interpreted as meaning that sufficient rater agreement exists to produce stable estimates of task mean values.

The standard REXALL analysis procedure for achieving acceptable rater agreement and a set of reliable task mean ratings is to eliminate divergent raters from the sample. Divergent raters are those raters whose ratings differ significantly from the ratings of good raters because of their deliberate non-cooperativeness in following instructions, inverted or poor discriminative use of the rating scale, unique perception of task, or lack of knowledge. These divergent rater characteristics are reflected by a low or negative correlation (Pearson r) between the individual rater's set of ratings and the sample task means (excluding the subject rater's ratings). A typical rater sample is assumed to have a simple structure consisting of a majority of good raters who formulate a set of stable task means and a minority of divergent raters who disagree with the majority rating pattern. For determining training emphasis, the rank-ordered task means computed from the ratings of the residual good raters constitute the recommended training priority and define the common rating policy (CRP).

REXALL analysis does not permit TE data displaying persistent low R_{11} and/or divergent raters after application of deletion procedures to be further processed. The rationale underlying the present research is that for such specialties, low R_{11} may be a function of multiple rating policies associated with sub-groups of raters sharing similar training perceptions aligned with specific employment areas within the specialty. If this is the case, mean ratings across a total specialty sample could conceivably dilute expert ratings on technically critical subsets of tasks to the point where they compete with less important general tasks for recognition in the final task training priority.

APPROACH

In establishing the research thrust, the following factors were initially regarded as possible causes for poor rater agreement (low R_{11}) in TE data: (a) sampling variations, (b) multiladder task lists, (c) random rater heterogeneity, (d) presence of divergent raters, and (e) multiple rating policies. The research reported in this presentation focuses on sampling variations by examining the effects of sample size on interrater reliability and examines the remaining possible causes of poor interrater agreement by assessing the results of two different analytical approaches. Overall, the research approach to investigating the multiple rating policy/poor rater agreement hypothesis was to employ two independent analysis techniques: (a) CODAP cluster analysis, and (b) factor analysis. A brief introductory outline for each technique is provided in the relevant findings section in this paper.

Sample size is an important consideration in the deliberation of possible causes for poor rater agreement. Average operational TE sample size is 45 raters with a range of 10 to 80 raters. Statistically, there is a greater chance of obtaining an unrepresentative sample with abnormally low (or high) rater agreement for the smaller samples. The relationship between sample size and the interrater reliability indices, R_{11} and R_{kk}, is algebraically summarized by the Spearman-Brown prophecy formula. In general terms it states that R_{kk} increases as R_{11} and sample size increase. The criterion minimum, $R_{11} = .20$, for acceptable rater agreement is obtained from this formula by insertion of $R_{kk} = .90$ as the widely recognized criterion minimum for stable task means, and a sample size of approximately 40 raters being regarded as sufficiently large to be stable. Estimation of this minimum safe sample size assumes the level of rater agreement and basis for agreement

(rating policy) within the sample reflects that of the parent population. To address the issue of the stability of R_{11} as a function of sample size, the two large single specialty samples were taken as independent finite populations; 100 sub-samples for each of 12 sample size points in the 10 to 100 rater range were randomly selected and assessed for level of rater agreement (R_{11}).

In the case of the multiladder condition where more than one specialty is surveyed with a single comprehensive survey instrument, low R_{11} would be attributed to conflicting specialty aligned interests with little or no common training recommended. REXALL analysis would obviously be inappropriate under this condition. Analysis of a dual specialty sample was included in the investigation of multiple rating policies both in combined form and as two single specialties.

The third factor acknowledges the possibility of a rater sample where most raters agree to disagree due to their highly individual interpretations of the task list and/or rating scale. This represents the extreme multiple rating policy condition with no meaningful applicable training recommendations. Although the research approach taken here uses cluster and factor analyses as primary methods, an understanding of how interrater agreement is assessed and how ratings policies are examined using existing techniques is in order. Being the only ratings analysis tool readily available in CODAP, REXALL is normally used for analyses of all ratings.

Standard REXALL analysis is based on the fourth cause, i.e., that the presence of divergent raters serves to depress sample rater agreement. Existing REXALL procedures for extracting a reliable CRP involve the deletion of the initial divergent rater set (pass 1) and, if necessary, deletion of any newly generated divergent rater(s) (pass 2). Consistently observed increases in R_{11} and R_{kk} resulting from the deletion of divergent raters in operational samples support this procedure and contribute to the face validity of the following operational CRP extraction criteria: (a) minimum acceptable level of rater agreement, R_{11} = .20, R_{kk} = .90; (b) rater divergency, $r <$ .30, (c) deletion confidence - maximum of two deletion passes, maximum of 10% raters deleted; and (d) desirable number of good raters, 40. Failure to achieve a reliable CRP via this procedure because of persistent low R_{11} and/or divergent raters results in specialties being considered complex. One possible interpretation of the complex rater sample is that it contains an inordinate number of divergent raters who disguise the underlying CRP to an extent which renders existing CRP extraction criteria unsuitable. However, considering the research to be driven by the quest for identifying multiple rating policies, the adequacy of these criteria was assumed.

Accepting the multiladder sample type as being obviously predisposed to being complex and unsuitable for REXALL analysis, the postulated single specialty rating policy domain is summarized in Figure 1. The simple or complex specialty classification corresponds to achievement or non-achievement of a reliable CRP employing the previously described standard REXALL analysis procedure/criteria.

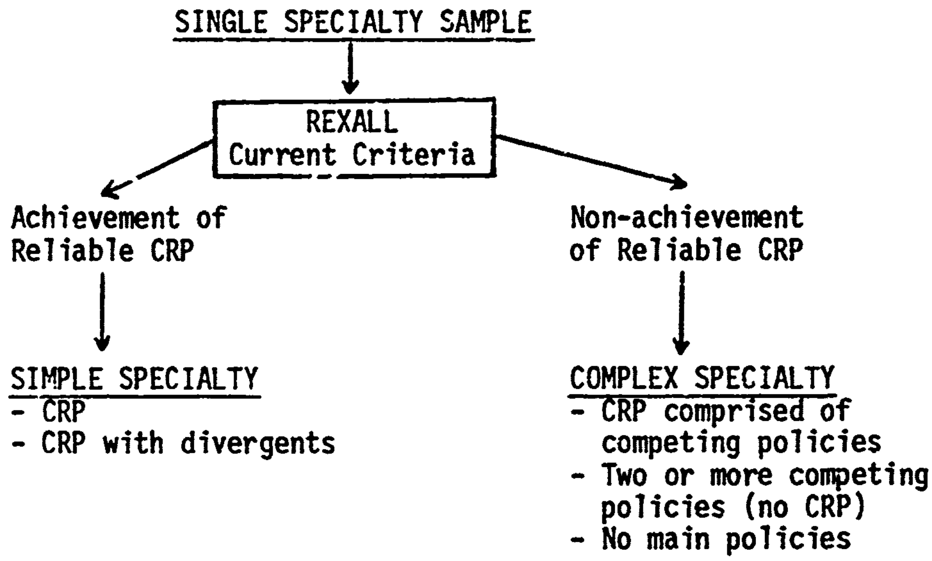


Figure 1. Single specialty rating policy domain

Multiple rating policies are defined in terms of differences in the rank-ordering of tasks between various subgroups of raters. A rank-order correlation $r_s < .50$ was taken as indicating a practical difference in the recommended training priority between any two rating policy groups. In relation to meaningful alternative training policies, it would be highly desirable for raters within significantly different rating policy groups to share a common background characteristic such as job title or major command (MAJCOM).

The analysis techniques were tested with TE data for five single specialties and a dual career-ladder which was analyzed both in the combined form and as two single specialties (see Table 1). All samples failed to qualify as a simple specialty (reliable CRP plus divergents) under strict application of the 10% deletion confidence criterion.

Table 1. Training Emphasis Samples Analyzed

AFSC	Title	Number of Raters
404X0	Precision Imagery and Audio-Visual Media Measurement	47
811X0	Security Specialist	120
328XX	Avionics Communications/Navigation Systems	148
328X0	Avionics Communications Systems	65
328X1	Avionics Navigation Systems	83
672X2	Disbursement Accounting	149
304X4	Ground Radio Communications Equipment	335

FINDINGS

The findings presented pertain to the research of sampling error and multiple rating policies as possible causes of poor rater agreement.

Sampling Variations

Table 2 details the variation in R_{11} ($\bar{X}$ & SD) at three sample sizes for the two specialties. The observed range in R_{11} (MIN, MAX) illustrates the extent to which observed agreement can differ from that of the parent population for a typical operational sample size in the 10 to 100 rater range. With respect to establishing a suitable sample size for REXALL analysis, both specialties are sufficiently stable at the 50 to 60 rater size to permit extraction of the CRP (if present). For sample sizes much below 50 raters, the problem of sampling error as a cause for poor rater agreement is more significant.

Table 2. Variation in R_{11} With Sample Size

Sample Size	R_{11} for AFSC 672X2				R_{11} for AFSC 304X4			
	$\bar{X}$	SD	MIN	MAX	$\bar{X}$	SD	MIN	MAX
10	.238	.112	.017	.517	.156	.061	.025	.205
50	.257	.033	.144	.335	.167	.020	.119	.214
100	.259	.021	.211	.308	.165	.012	.132	.196
N	R_{11}=.2596				R_{11}=.1686			

Detecting Multiple Rating Policies

Cluster Analysis

The CODAP clustering programs were applied to the samples in an attempt to develop new procedures and guidelines for using and interpreting existing occupational clustering software with task factor data. For all samples the percent training emphasis overlap algorithm aggregated raters who were very homogeneous with respect to the number and type (by duty) of tasks rated. REXALL analysis of these main rater groups produced significantly higher values of R_{11} than observed with the parent sample, indicating that once raters are found to have high overlap with one another on the ratings of tasks they choose to recommend for training, they have a high level of rater agreement

The following limitations are seen as major obstacles to accepting the training emphasis cluster structures as a suitable method for defining multiple rating policies: (a) the requirement to adjust ratings to a percentage of a rater's total rating sum results in the loss of important

information about the level (magnitude) of assigned ratings, (b) the overall clustering is strongly driven by overlap across all (non-zero) rated tasks which detracts from common duty emphases, (c) subjective decisions are required to determine the cluster group boundaries, and (d) the status of the considerable number of isolate raters (5%-20%) is an unknown.

Factor Analysis

A Q-Type principal components factor analysis was applied to each sample. With this approach, raters were treated as variables loading on factors (dimensions of common variance) which were interpreted as potential rating policies. The customary criterion loading of .33 was taken as the minimum absolute value for meaningful rater contribution to a factor rating policy. In contrast to cluster analysis, where rating policies are characteristic of rater groups with mutually exclusive membership, factor analysis generates rating policies which are external to the rater set by determining each rater's loading on each rating policy extracted. This permits evaluation of rater performance across all policies. A further feature of this approach is the capability to control the number of rating policies for analysis. Initially, the extent to which a single general factor common rating policy prevails was investigated. By employing a VARI-MAX rotation/factor building methodology the relative utility of factor solutions consisting of iteratively increasing numbers of rating policies was evaluated.

General factor solution. The general factor extracted in a one-factor solution accounts for the greatest amount of shared variance within the data, and is conceptualized as the CRP underlying the total rater set. Analysis of the pattern of rater loadings on this factor establishes the extent to which the CRP exists within the sample. All single specialty samples were found to have a factor CRP characterized by: (a) all significant loadings being unidirectional, and (b) an acceptable level of rater agreement. For all but the least agreeable sample (AFSC 404X0) these factor CRPs accounted for the majority of raters. In contrast, the dual specialty (AFSC 328XX) general factor was comprised of bipolar significant loadings indicative of two strongly opposing specialty-specific rating policies and preclusive of a CRP as the dominant policy for the total sample. For all single specialties, iterative removal of raters from the low loading end of the rank-ordered general factor loading sequence resulted in a steady increase in R_{11} and R_{kk}, establishing this sequence as an accurate distribution of rater performance with respect to the CRP. Comparison of the REXALL high-low rater correlation sequence (as produced by the sample mean vector) with the corresponding general factor high-low rater loading sequence for each single specialty revealed a close matching in rater rank orders and correlation/loading values which tended to virtual equivalence with increasing total sample R_{11}. Except for AFSC 404X0 the REXALL grand task mean vector performed adequately as a standard for determining the relative worth of all raters with respect to the CRP.

The information conveyed by the one factor solution, together with the factor/REXALL analyses comparisons, permit modification of the original REXALL CRP extraction criteria described in the report background. In general terms, these findings demonstrate for the single specialty samples, the reliable CRP is derived via REXALL analysis when a level $R_{11} \geq .20$ and $R_{kk} \geq .90$ is attained by the successive deletion of sets of divergent raters ($r < .30$),

providing R_{11} increases with each deletion pass and no more than 25% to 30% of the sample is deleted. Allowing for the deletion of this maximum number of divergent raters, and taking into account the R_{11} stability/sample size findings, it was found that a minimum sample size of 55 raters was required to attain minimum acceptable rater agreement. For smaller samples dictated by rater availability, $R_{11} \geq .20$ and $R_{kk} \geq .80$ would be acceptable.

Rotated factor [illegible]ons. Application of the VARI-MAX rotation/factor building technique [illegible] samples identified different rating policies ($r_s <$ .50) in two ins[illegible] complex single specialty, AFSC 404X0; and the dual specialty sample, [illegible] 328XX. For all other samples the rotated solution analyses reinforced the CRP as the dominant rating policy by identifying two or three main internal rating themes as minor variations of the CRP.

The competing multiple rating policies within AFSC 404X0 and AFSC 328XX render these total samples complex and unsuitable for REXALL analysis. The rotated solutions for the remaining five single specialty samples share common features which disqualify the component factors as meaningful multiple rating policies. These five single specialties are appropriately classified as simple or non-complex in that the REXALL CRP reliably subsumes the competing component rating policies.

CONCLUSIONS

1. Factor analyses of the single specialty training emphasis samples in this report have demonstrated them to be less "complex" than anticipated.

2. REXALL analysis employing the new CRP extraction criteria is adequate for the following sample types: (a) CRP with no divergent raters (ideal); (b) CRP with divergent raters, and (c) CRP comprised of competing rating policies.

3. Rexall analysis is inadequate for the following sample types: (a) two or more competing rating policies, eg, AFSC 404X0; (b) no rating policies; and (c) multiladder surveys, eg, AFSC 328XX

4. CODAP cluster analysis is not adequate for identifying multiple rating policies.

5. Principal component factor analysis has a high utility for identifying the CRP and multiple rating policies.

A COMPARISON OF TWO BASIC RUSSIAN TEACHING METHODOLOGIES [DLI's 47-Week Basic Course and the NCS RS-15A Program]

Ronald C. Jenkins, Department of Defense
Lee Corrigan, Defense Language Institute, Foreign Language Center

I. INTRODUCTION

This report contains the results of a comparison of two basic Russian language teaching methodologies -- the RS-15A program as presented at the National Cryptologic School (NCS), which is a contextual method with multiple teaching strategies, and the Basic Russian Language Course, a multiple-strategy approach with audio-lingual overtones, as presented by the Defense Language Institute, Foreign Language Center (DLIFLC), Monterey, California.

II. BACKGROUND

The RS-15A Russian curriculum is the core element of the Russian Linguist Acquisition Program (RLAP), a recruitment and training program developed at the NCS, and designed to produce competent Russian linguists through academic and on-the-job training.

The Defense Language Institute Foreign Language Center's Russian basic language curriculum relies heavily on rote memory by the student of word patterns accompanied by lessons in the grammar aspects of the language. All the DLIFLC instructors employed in the curriculum are native speakers of the foreign language being taught. They may or may not be proficient English speakers. They emphasize proper pronunciation and word patterns presented in daily memorized dialogs with the underlying philosophy that the student will recall the dialogs and be able to function efficiently in the foreign language should the need occur.

The RS-15A Russian curriculum was modified slightly to control the time variable. (The original RLAP program was 42 weeks of class work followed by an on-the-job training period.) The class time was increased to 47 weeks to be equal with the DLIFLC training period and because the students selected for the military training program were not subject to the rigorous RLAP screening process.

III. THE RATIONALE FOR THE STUDY

While the DLIFLC methodology has served foreign language instruction well for several decades, recent research conducted separately and cooperatively by government agencies has indicated that audio-lingual techniques which concentrate on pattern drills without proper attention to the contextual aspects of the language may not be as effective in teaching the deeper structures of the language as a method that incorporates study of the grammar, syntax, and contextual features.

Since the RS-15A Russian curriculum incorporates grammar instruction with pattern drills along with syntax and cultural studies, it should (and under controlled conditions using highly selected students has) produce linguists with measurably higher proficiency than methods where such instruction was not used.

IV. THE METHODOLOGY OF THE STUDY

Three groups -- one at the NCS and two at DLIFLC -- were formed. The NCS group, as stated earlier was instructed in the RS-15A Russian curriculum material and methodology while the two groups at DLIFLC received instruction in the standard DLIFLC course material and methodology. The reason for the two groups at DLIFLC was to check for a "Hawthorne" effect.

One variable that could not be foreseen was the previous foreign language studies (in languages other than Russian) of the NCS Group. This group had twice the experience in foreign languages as the other groups.

For the purposes of this experiment, measurements of attained foreign language proficiency were made using group mean scores and significance was determined at the .05 level.

Since the only criterion for selecting candidates for the experiment was that they meet standard DLIFLC entrance requirements, no effort was made to influence the composition of either group.

V. THE EVALUATION PLAN

Foreign language learning aptitude was determined through the administration of the Defense Language Aptitude Battery (DLAB).

While it is yet uncertain just what role knowledge of the grammar of one's mother tongue plays in foreign language acquisition as an adult (studies being conducted at DLIFLC are as yet inconclusive), the English Grammar Recognition Test (EGRT) was administered to all students during the first week. The data was used to determine differences in English skills among the three groups.

In addition to measuring the outcome of the process, it was desirable to assure that the process was properly conducted in order to increase the confidence that the outcome, whether the results were favorable or not, was a result of the process. To this end, a Student Opinion Form (SOF) was developed to gather data on the conduct of the students, the instructors, and the course operations. The SOF was administered twice during the 47 weeks of study, in the 16th and 46th weeks.

The following tests were administered at the end of course to determine the proficiency of the students.

a. Defense Language Proficiency Test (DLPT) -- A multiple choice, paper/pencil and magnetic tape, norm-referenced test which tests the candidate's ability to understand the spoken language and to read the language.

b. Standard Proficiency, Entry Level 2 Test (SPEL-2) -- A contextual reading test designed to evaluate the candidate's understanding of the foreign language syntax and semantic structure at language level 2.

c. Language Proficiency Test (LPT) -- A contextual reading test designed to operate the same as the SPEL-2 test but which tests at language level 2+.

VI. ANALYSIS OF THE DATA

The following analysis of the data was performed to maximize the value of the primary findings, investigate the operations of the proficiency tests.

1. A one-way analysis of variance (ANOVA) of the EGRT scores to insure that there were no significant differences among the groups in English skills.

2. A one-way ANOVA of the DLAB scores to insure there were no significant differences in aptitude among groups.

4. A one-way ANOVA of the DLPT scores collected at the mid-point of the course for DLIFLC Group A and the NCS Group to determine if significant differences among groups could be determined at that point.

VII. CONDUCT OF THE EXPERIMENT

Although the experiment ran smoothly for the most part, all groups suffered attrition. In addition, due to scheduling problems, not all the DLIFLC subjects were tested with all of the tests during the final week. In the interest of including as many subjects as possible in the experiment, subjects with incomplete test data were included but allowances were made for missing data points during the statistical calculations. For the final analysis of results, DLIFLC Group A comprised 18 subjects; DLIFLC Group B, 15 subjects; and the NCS Group, 17 subjects.

VIII. FINDINGS

The ANOVA's were performed on DLAB and EGRT scores and the results showed that the students in the three classes were from the same population. Further, the ANOVA's indicated that the DLIFLC classes operated as one; there was no "Hawthorne" effect.

An analysis of the SOF surveys indicated that the students at the two locations viewed their instructors as effective and capable and that the students' study habits were similiar. The SOF did indicate that the DLIFLC students were required to perform more military duties than their NCS counterparts and this probably impacted the final measured language proficiency.

The language proficiency tests were administered as scheduled and ANOVA's run for the scores. The DLPT failed to measure any significant differences among the groups but the LPT and the SPEL-2 tests measured significant differences.

IX. TEST RESULTS CONVERTED TO STANDARD SCORES

To illustrate the measured differences among the groups on the language proficiency tests, the group mean scores have been converted to Z scores and plotted in the tables below.

Table 1
Z SCORES FOR THE LPT

Group	Part 1	Part 2
DLIFLC A	-.70	-.53
DLIFLC B	.04	-.20
DLIFLC A & B	-.32	-.36
NCS	.47	.53

Table 2
Z SCORES FOR PART 1
OF THE LPT PLOTTED

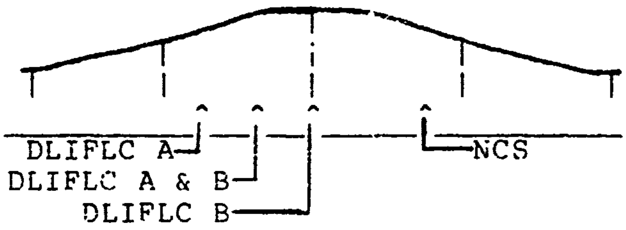
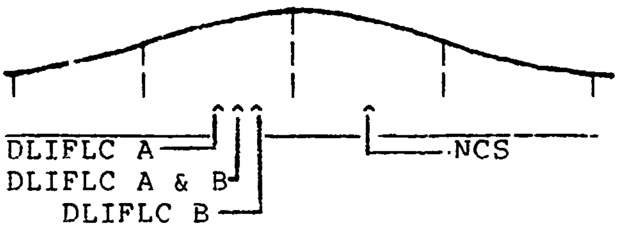


Table 3
Z SCORES FOR PART 2
OF THE LPT PLOTTED



For Part 1 of the LPT, the NCS Group scored:
- -- 1.17 standard deviations higher than DLIFLC Group A
- -- .43 standard deviation higher than DLIFLC Group B
- -- .79 standard deviation higher than the DLIFLC Groups A & B

For Part 2 of the LPT, the NCS Group scored:
- -- 1.06 standard deviations higher than the DLIFLC Group A
- -- .73 standard deviation higher than DLIFLC Group B
- -- .89 standard deviation higher than DLIFLC Groups A & B

Similar results were obtained using the SPEL-2 language proficiency test.

Table 4
Z SCORES FOR THE SPEL-2 TEST

Group	Part 1	Part 2	Total
DLIFLC A	-.19	-.23	-.21
DLIFLC B	.02	-.70	-.48
DLIFLC A & B	-.07	-.44	-.33
NCS	.14	.86	.64

Table 5
Z SCORES FOR PART 1 OF
THE SPEL-2 TEST PLOTTED

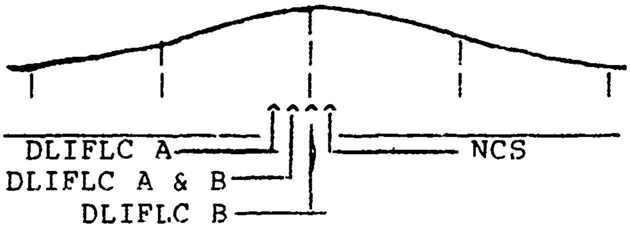
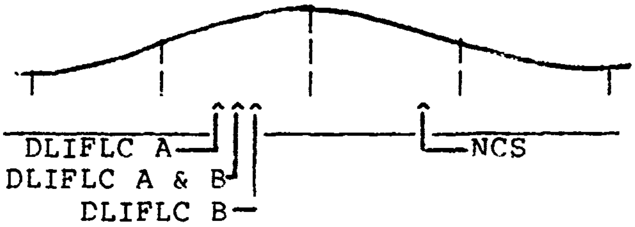


Table 6
Z SCORES FOR PART 2 OF
THE SPEL-2 TEST PLOTTED

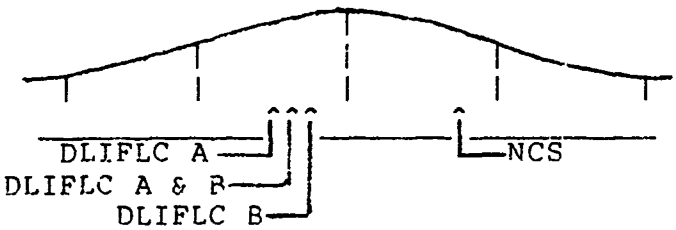


For Part 1 of the SPEL-2 Test, the NCS Group Scored:
- -- .25 standard deviation higher than DLIFLC Group A
- -- .16 standard deviation higher than DLIFLC Group B
- -- .21 standard deviation higher than DLIFLC Groups A & B

While the performance of the NCS Group is higher than that of the DLIFLC students on part 1 of the NCS SPEL-2 test, the ANOVA discussed earlier showed that the differences are not significant at the .05 level.

For Part 2 of the SPEL-2 Test, the NCS Group scored:
- -- 1.09 standard deviations higher than DLIFLC Group A
- -- 1.56 standard deviations higher than DLIFLC Group B
- -- 1.30 standard deviations higher than DLIFLC Groups A & B

Table 7
TOTAL SCORE FOR THE SPEL-2 TEST PLOTTED



Overall, on the SPEL-2 Test, the NCS Group scored:
- -- .85 standard deviation higher than DLIFLC Group A
- -- 1.12 standard deviations higher than DLIFLC Group B
- -- .97 standard deviation higher than the DLIFLC Groups A & B

Table 8
Z SCORES FOR THE DLPT

Group	Part 1	Part 2
DLIFLC A	.06	-.14
DLIFLC B	-.09	-.05
DLIFLC A & B	-.01	-.10
NCS GROUP	.01	.19

Table 9
Z SCORES FOR PART 1 OF THE DLPT

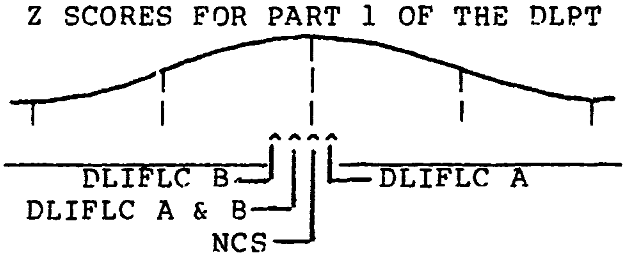
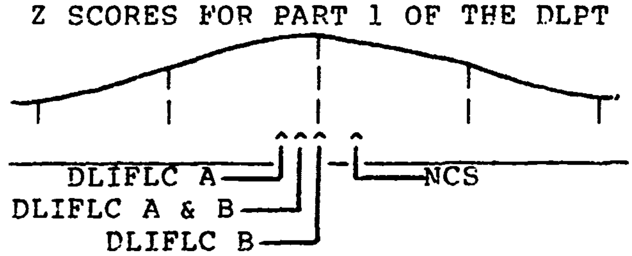


Table 10
Z SCORES FOR PART 1 OF THE DLPT



For Part 1 of the DLPT, the DLIFLC Group A
-- .05 standard deviation higher than the NCS Group
-- .07 standard deviation higher than DLIFLC Groups A & B
-- .15 standard deviation higher than the DLIFLC Group B

For Part 2 of the DLPT, the NCS Group
-- .33 standard deviation higher than DLIFLC Group A
-- .24 standard deviation higher than DLIFLC Group B
-- .29 standard deviation higher than DLIFLC Groups A & B

X. DISCUSSION AND CONCLUSION

The comparison of the NCS RS-15A curriculum with the DLIFLC Basic Russian curriculum was conducted as much as possible along classic experimental lines. The three groups began the experiment on an equal footing and the curricula was presented by the instructors as designed. The students at the two locations behaved similiarly in study habits, etc. There was one unpredicted difference that almost certainly impacted the outcome of the experiment; the DLIFLC Group was required to perform more military duties than the NCS Group.

Only two of the final tests used measured any differences among the groups. The measured differences, however, were significant at the .05 level, with the NCS Group achieving higher levels of proficiency in all instances.

Although there were two variables that were not controlled in the experiment which probably affected its outcome, the significant measured differences in language proficiency as tested by two of the three language proficiency tests in the experiment strongly support the hypothesis that the RS-15A cur-curriculum can produce more competent linguists than the course used by DLIFLC at the time of the experiment.

PROVIDING THE MAN IN THE BACK SEAT

by ALAN JONES
Department of Senior Psychologist (Naval), United Kingdom

INTRODUCTION

Despite his passive-sounding name the Royal Navy Observer is a very active and important member of a helicopter crew. He acts as a navigator and tactical coordinator in 2- and 4-man helicopters. He operates variable depth sonar, radar, electronic warfare equipment, and weapons systems such as homing torpedoes, depth charges, and air to sea missiles.

The Royal Navy has met various problems in recruiting suitable personnel, that is, in providing this important man in the back seat. The majority of Observers join the Service on a Short or Medium Career engagement, and only a relatively small number come from amongst Full Career Officers. The Short/Medium Career Commission for aircrew in fact provides the majority of both Pilots and Observers in the Royal Navy.

For Short/Medium Career Officers the training pattern until recently has been 8 months basic Officer training, twenty four weeks Basic Flying Training (100 flying hours in a fixed wing aircraft), Advanced Flying Training (27 flying hours in a helicopter), followed by 50 flying hours of Operational Flying Training. Over a seven year period Short/Medium Commission Officers have shown a 28% wastage rate during Officer training, of which half was voluntary. During Basic Flying Training (BFT) there was 29% wastage (40% of those entering), and 7% (16% of those entering) at Advanced Flying Training (AFT). The overall wastage has therefore been approximately two-thirds of those recruited.

From the above figures two major problem areas can easily be identified: voluntary wastage in Officer training and wastage in Basic Flying Training. Failure at this latter stage is predominantly (90%) ascribed to problems in the air (as distinct from ground school). Often trainees are described as "leaving some of their brains behind on the ground."

RECRUITMENT AND SELECTION OF OBSERVERS

Faced with such a situation there are a number of (theoretically) straightforward approaches to study and possibly to rectify the situation which a psychologist can adopt. One obvious step is to examine the existing selection system and procedures and carry out appropriate predictive validity research.

The two main stages in the procedure are the aptitude tests (for aircrew potential) and the assessment centre (for Officer potential). The aptitude tests are those used by the Royal Air Force for Pilot and Navigator selection and are basically updated versions of tests produced by the end of World War 2. From the tests an index of suitability is produced, using different weights for the various tests.

Examination of the predictive validity of this index and of the individual tests has been complicated by factors such as relatively small sample

sizes, changes in the training system, the effect of the voluntary wastage rate and so on. However it soon became clear that the index had deficiencies. In fact one of the 5 tests used (Mathematics) was providing most of the predictive power. The weighting of the index has subsequently been changed, but it appears unlikely that further reweighting of existing tests will give much improvement.

Although the aptitude tests might be expected to predict training performance, it is probably unreasonable to expect them to predict voluntary withdrawal. However, market research and attitude survey results may help clarify the broad reasons for withdrawal.

One target market research study and one attitude survey of aircrew have been carried out. The most important result from the market research study was that first and foremost aircrew applicants are committed fliers (rather than Naval Officers) and have often thought of the RN rather late in the process of career choice. Since the prime motivation is to fly, it is not surprising that Observer is very much a second choice; a number of those surveyed said they would accept it if it was the only way to fly, whilst some said that they would not accept it on any terms. Few candidates have even the vaguest notion of what an Observer does.

The attitude survey (of serving and ex-aircrew) confirmed that helicopter flying appears to attract a large number of young men who are outside the standard catchment area for Armed Service Officer recruitment. Again, despite all the efforts made in the literature and elsewhere, some entrants thought it was difficult to find out what an Observer actually does and putting Observer as second choice may be done more to look enthusiastic about a Naval career than as a result of a genuine interest in the Observer role. In fact, even after all the recruiting literature and counselling given, the number putting Observer as first choice is still small (around 1%).

The difficulty of attracting Observers (which really means attracting Observers from those who very strongly wish to be Pilots) has led to more recruitment resources being deployed in this area. However, it may be that, as well as attempting this, it would be sensible to aim at non-fliers, either by allocating more Full Career Officers to this specialisation or by attracting Short Career Seaman Officers. Both these groups are likely to suffer less from the "frustrated pilot syndrome" and may have acquired relevant skills from experience in surface ships.

Whilst the studies quoted so far have suggested ways in which recruitment and initial selection might be improved, it appeared unlikely that they alone would solve the problem. When the first test validation results become available, there were doubts about whether the process of cross-validation of a reweighted index or the development of new aptitude tests would be likely to achieve any short- or medium-term solution. There was also the consideration of sustaining Observer motivation and developing knowledge of the Observer role during Officer training. Accordingly emphasis was put on the development of a trainability test (known as "miniaturised training tests" in the USA).

TRAINABILITY TESTS

Trainability tests, as their name suggests, are aimed at predicting training

success (rather than job performance). The selector is attempting to predict the extent to which the individual will successfully cope with training. Some form of teaching and learning is included in the predictor measure so that a sample of training behaviour can be observed and used as a basis for prediction. A structured and controlled learning period is therefore an important aspect of the testing procedure. Care must be taken that this preparatory procedure incorporates existing instructional and training approaches. During the test itself how things are done are observed as well as what is done.

Assessment is usually made by instructor-assessors, using an error checklist and giving an overall rating (grade) of the likelihood of completing training. Robertson and Downs (1979) cite validity evidence for a number of training courses (jobs such as carpenter, fork-lift truck driver, and sewing machinist). From 16 predictive validity studies (total N = 835), the average validity coefficient was .44 (range .02 to .72) for the error checklist and .54 (range .04 to .81) for the assessor's rating. These validity coefficients are encouraging and, as Reilly and Chao (1982) state, are at least comparable with validities for standardixed tests and biodata. They appear most valid for short (6 months or less) training courses. Because of the method of test development they also have high content and face validity.

At an early stage in test development the task to be used as the test has to be determined by analysis of the training course, usually using a critical incidents technique. The task has to be based on crucial elements of the job, use only such skill and knowledge as can be imparted during the learning period, be sufficiently complex to allow a range of observable errors to be made, and be capable of being carried out in a reasonable time.

The main benefits of trainability tests are high predictive, face, and content validity, but their main disadvantage is cost, since trained personnel, equipment and materials are involved.

THE OBSERVER TRAINABILITY TEST

Because of the high failure rate (40% of those entering) in Basic Flying Training (24 weeks of 100 flying hours) it was decided, after an initial feasibility study, to develop a trainability test (or "Grading") aimed at predicting success in this stage of training.

It should be noted that up to this point the trainability test technique had been applied to jobs with a psychomotor skill emphasis. This was the first time a high-level, essentially cognitive job had been tackled.

From an analysis of problems encountered by trainees in BFT, it was found that trainees tended to have problems in navigation and radar sorties, with ground exercises and tactical navigation less of a problem. Eight main types of error were identified, for example poor pre-planning, memory of procedures, and accuracy of computation.

The analysis we carried out emphasised the difference between the air and ground situation. An individual who copes well on the ground might perform badly in the air. This may be because of the paced effect of being in an aircraft, the realisation of the actual responsibility in directing the aircraft, or the constantly changing signals. It was felt that an airborne trainability test offered the most feasible way of assessing the likelihood

that a trainee could cope with the flying aspects of training. However, exercises are also carried out on the ground; the same dead reckoning navigation exercise is given twice and subsequent results have suggested that performance here has some predictive value.

The test is carried out at the air station where BFT takes place. Students receive 14 hours ground instruction and receive a Student Study Guide to help them prepare for the classroom work and the trainability test itself. Because the test takes place toward the end of Officer training they will already have had some experience of some relevant areas, particularly of navigation. The trainability test instruction covers areas such as direction, air speeds, velocity, use of Dalton computer, radar modes, fixing, plotting and RT procedures. A briefing is also given on the trainability test flight itself, and time is allowed for familiarisation with instrumentation.

The flight lasts 1 hour 30 minutes. One student is graded per flight by an experienced Observer instructor, whose main job is Grading and who is not involved in BFT. He uses an error checklist (with 55 points overall), completes 6 ratings (e.g. on airsickness), marks the student's chart and log card (using the system used in BFT itself), and makes an overall assessment of the student's likelihood of getting through BFT. A six grade system is used:

GRADE	DEFINITION	% CHANCE OF PASSING BFT
A	This student should do well in basic training	90-100
B	This student should pass basic training	70- 89
C	This is a marginal student	50- 69
D	This student is likely to be a training risk but might get through	40- 49
E	This is a poor student who is likely to fail training	25- 39
F	This student can be expected to fail basic training.	0- 24

The flight is in a Jetstream aircraft (cruising speed around 200 mph), at heights between 500 and 2000 feet, and is as follows. After take-off the instructor directs to an on-top position off a small island and then carries out one modified curve of pursuit homing to a ship contact 24 to 30 nautical miles away as a demonstration. The student then directs the aircraft back to the island, advised throughout by the Instructor. When at the island, a debrief given if required. The student then carries out 2 modified curve of pursuit homings - one to a target at a range of 24 to 30 nautical miles and then one back to the island. The aircraft then climbs. The gradee than passes the heading for the first point on a 2-leg navigation exercise (using dead reckoning and navigation aids). In fact the course is essentially that of the dead reckoning tests carried out earlier on the ground, but data must be read off the instrumentation.

During this section he is assessed on his ability to navigate. The aircraft then returns to base and the examiner makes the overall assessment based on

the error checklists, objective scoring of the log and of chart, and his impressions of the gradee's performance. The pilot also contributes a rating on the voice communication from the student to direct the aircraft.

Little work has been done on the inter-rater reliability of trainability test assessments but there are reasons for believing it is high (Robertson and Downs, 1979). In the case of the Observer trainability test only one assessor is appointed to this job at a time and so the problem is one of ensuring an efficient handover and maintenance of the current standards. So far, over the 2 years or so of grading, 3 assessors have been involved and care has been taken in the handover period. Analysis of their ratings has not shown any individual leniency or harshness effects. Nor has there been any evidence of any effect of the order in which individuals in the group are flown and graded. So far 63 direct entry Short/Medium Career Observer trainees have been graded, giving the distribution of overall grade shown in Table 1. Roughly 20% have been considered relatively unlikely to get through training (grades E and F), and it would be individuals so assessed who would probably be excluded if the trainability test were used executively. Results have been kept from trainers and trainees to avoid any possibility of self-fulfilling prophesies.

TABLE 1. DISTRIBUTION OF TRAINABILITY TEST GRADES

GRADE	DEFINITION - % CHANCE OF PASSING BASIC FLYING TRAINING	PERCENTAGE FREQUENCY
A	90-100	5
B	70- 89	27
C	50- 69	29
D	40- 59	20
E	25- 39	14
F	0- 24	5

Of those graded, 43 have entered flying training and should have completed Basic Flying Training. The relationship between grade and failure in training is shown in Table 2. There is a progression of percentage failure as one moves down the grades, with the observed frequencies reasonably in line with the definitions for each grade (see above). The biserial r correlation given by the data is .48.

The two grade Es who completed BFT are worthy of brief examination. One withdrew voluntarily at the beginning of the next stage of training and so is difficult to consider a real error of prediction. The other individual (assessed early on in the project) appears to have been misclassified by the assessor as his recorded performance during the test might be expected to yield a C or a D. This one case perhaps illustrates the difficulty

of using observations made during the Grading flight, where subjectivity may unduly intrude.

TABLE 2. TRAINABILITY TEST GRADE AND BASIC FLYING TRAINING

GRADE	N	%FAIL BFT
A	2	0
B	14	22
C	11	45
D	7	57
E	8	75
F	1	100

A correlation of .48 (even if a little higher in the light of the above comments) is a little lower than might be expected on the basis of previous trainability test research. One reason for this may be that for previous jobs studied, it was easier to observe the students performance (eg carpentry). With the Observer the tasks are more cognitive and it may be more difficult to "see" what the individual is doing (besides such gross behaviours as airsickness); the error checklist may therefore be less effective in this situation.

DISCUSSION AND CONCLUSION

This paper has attempted to describe some approaches to studying and remedying a costly situation. As such it perhaps exemplifies the view that no one methodology will solve a problem; a combination is required. It also has detailed a methodology not so well known in the USA as in the UK, and its application to a high-level job.

One reason for the initiative in trainability testing was a belief that, in the short term, it offered more chance of success than the development of aptitude tests, but it may be that, particularly with mini- and micro-computers, tests which capture the more dynamic elements of the Observer's job can now be considered. Another important reason was the need to improve the perceived status of the Observer and to improve trainees' knowledge of the Observer role.

So far the idea of the trainability test has been well-received by the training authorities and by trainees. Interim validation results are encouraging and it is hoped that it will soon make a cost-effective contribution to providing the important "man in the back seat."

ACKNOWLEDGEMENT

The development of the trainability test was carried out by SYLVIA DOWNS (then of the Industrial Training Research Unit, now of the University of Wales Institute of Science of Technology), JOHN HODGKISS of Senior Psychologist (Naval), and Lieutenant Commander BRINSLEY SHERIDAN, RN Retd (then of the Observer School, RNAS Culdrose).

REFERENCES

1. REILLY, R R and CHAO, G T Validity and fairness of some alternative employee selection procedures. Personnel Psychology, 1982, 35, 1-62.

2. ROBERTSON, I and DOWNS, S Learning and the prediction of performance. Development of trainability testing in the United Kingdom. Journal of Applied Psycholcgy, 1979, 64, 42-50.

Comparison of the Coast Guard's Selection & Classification Batteries

Karen N. Jones

U. S. Coast Guard Institute, Oklahoma City, Oklahoma

This report compares examinee performance on the selection and classification batteries used by the U. S. Coast Guard -- the Coast Guard Selection Test (CGST) and the Navy's Basic Test Battery (BTB). For this investigation, test information curves were used to compare location of optimal measurement and measurement accuracy (or quantity of information) on the operational form of the CGST, an alternate form of the CGST, and Form 6 of the BTB.

Description and History of Batteries

Each test battery analyzed in this report contains three tests: General Classification Test (GCT) or Verbal Performance Test (VPT) with approximately 65% verbal analogy items and 35% word usage items; Arithmetic Test (ARI) with approximately 40% computation items and 60% reasoning items; and Mechanical Test (MEC) with 100% mechanical comprehension items. Additional information on the three batteries is presented below.

Operational Form of the CGST. To determine if applicants meet the Coast Guard's minimum mental requirements for enlistment, the Coast Guard administers a short (one hour) selection test at its recruiting stations. In 1979 the CGST replaced the Navy's Short Basic Test Battery (SBTB) as the Coast Guard's primary selection test. The operational form of the CGST contains 110 items: 50 on the GCT, 35 on the ARI, and 25 on the MEC. To develop the CGST, the Coast Guard revised items from obsolete forms of the SBTB and BTB, administered the items to Coast Guard applicants for enlistment to collect item statistics, and selected items for inclusion in the battery. The CGST was standardized against the BTB at recruit training centers. In March 1980, the Coast Guard extended the time limits on the CGST to ensure that all of the tests were power tests for approximately 90% of the examinees. In this paper, the data from the operational form of the CGST with the original time limits and with the revised time limits were analyzed together.

Alternate Form of the CGST. The alternate form of the CGST compared in this report is one of two alternate forms which were developed by the Coast Guard. This form has 105 items: 45 on VPT, 35 on ARI, and 25 on MEC. It was developed using items from the operational form of the CGST, items from a Coast Guard item-writing contract, and items developed in-house. To develop the test, items identified as potentially biased (Scheuneman, 1979) were excluded from the item pool and items providing high information, as measured by the item information curve in the range on the ability continuum near the pass/fail score used by the Coast Guard, were selected for the test. To locate the desired range on the ability continuum, the pass/fail Navy Standard Score (NSS) on the CGST was divided by 3 to arrive at a NSS of 45 per test. This NSS score of 45 was increased to a NSS range of 44-46. Using the conversion tables and test characteristic curves for the operational form of the CGST, this NSS range was converted to an ability range on each test. The alternte form of the CGST was administered at recruiting stations in March-May 1982 to collect data for standardizing the alternate form of the CGST against the operational form of the CGST.

Form 6 of the BTB. The Coast Guard administers a classification battery to all recruits at its recruit training centers to determine each recruit's eligibility for training

at "A" schools. The Coast Guard uses Form 6 of the BTB, which was developed by the U. S. Navy, as its classification battery. The three tests from the BTB analyzed in this study are: GCT with 100 items, ARI with 50 items, and MEC with 50 items. A description of the development and standardization of these three BTB tests with examples of the types of items included in each test is presented in Rimland (1958a and 1958b) and Swanson (1958). With the exception of the MEC, the tests and conversion tables used by the Coast Guard are the same as the ones originally developed by the Navy. The MEC developed by the Navy contains 50 tool knowledge and 50 mechanical comprehension items. In May 1981, the Coast Guard began using the MEC without the tool knowledge items. The Coast Guard standardized the 50-item mechanical comprehension part of the MEC against the total MEC (50 tool knowledge items and 50 mechanical comprehension items) on the BTB. In this study, the data from only the mechanical comprehension part of the MEC were analyzed.

Procedure

Data. Data from the following examinees were used in this investigation: (1) approximately 5,600 recurits who took the BTB at the Coast Guard's recruit training centers in 1980 and who had entered the Coast Guard by taking the CGST and (2) applicants who took the operational CGST and one of the tests on the alternate form of the CGST as part of the enlistment screening procedure in March-May 1982 (700-900 examinees per test on the alternate form of the CGST.)

Analysis. To compare the selection and classification test batteries, the three-parameter logistic model was selected. In this item response theory (IRT) model, the item parameters are: item discrimination, a; item difficulty, b; and lower asymptote, c (Hambleton and Cook, 1977). The item parameter estimation program ANCILLES (Urry, 1975) was used to obtain item parameters for each test (e.g., GCT on the BTB). Since the IRT item parameters differ in origin and unit of measurement across sets of items and test administrations, it was necessary to transform the items on equivalent tests (e.g., items from the ARI on the three batteries) to the same metric. To accomplish this, items from equivalent tests on the BTB and on the operational form of the CGST were calibrated together as one test. The items from equivalent tests on the operational and alternate forms of the CGST were also calibrated as one test. Using the items on the operational form of the CGST as a "linking test," the items on the alternate form of the CGST were transformed to the same metric as the items on the BTB and the operational form of the CGST (Marco, 1977).

Using these item parameters, the test information curves were obtained for each test on the operational form of the CGST, the alternate form of the CGST, and the BTB. Since the number of items per equivalent test varied across batteries and the test information curve is a summation of the information provided by all of the items, a problem was encountered in comparing the test information curves. Therefore, the quantity of information provided by each test was multiplied by a constant to transform the test information curves for equivalent tests to the same scale.

Results and Discussion

The test information curves for the GCT/VPT, ARI, and MEC are presented in Figures 1, 2, and 3, respectively. Each figure contains the test information curve for the operational form of the CGST, for the alternate form of the CGST, and for Form 6 of the BTB.

Comparison of the BTB and the Operational Form of the CGST. As a review of these figures indicates, the most noticeable difference across test batteries is the fact that the

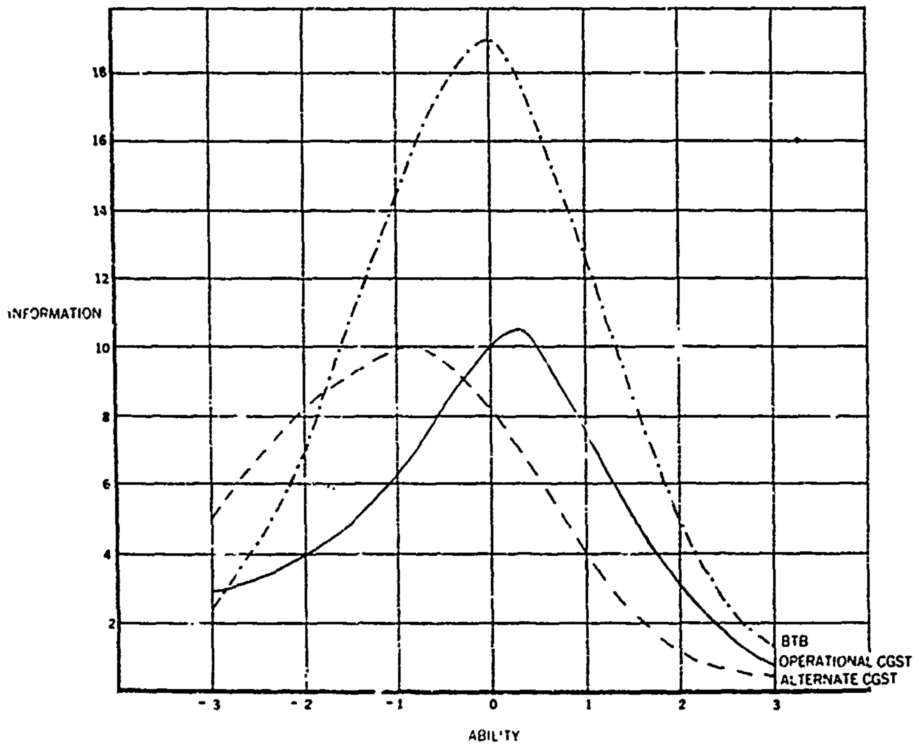


Figure 1. - Test information curves for GCT and VPT.

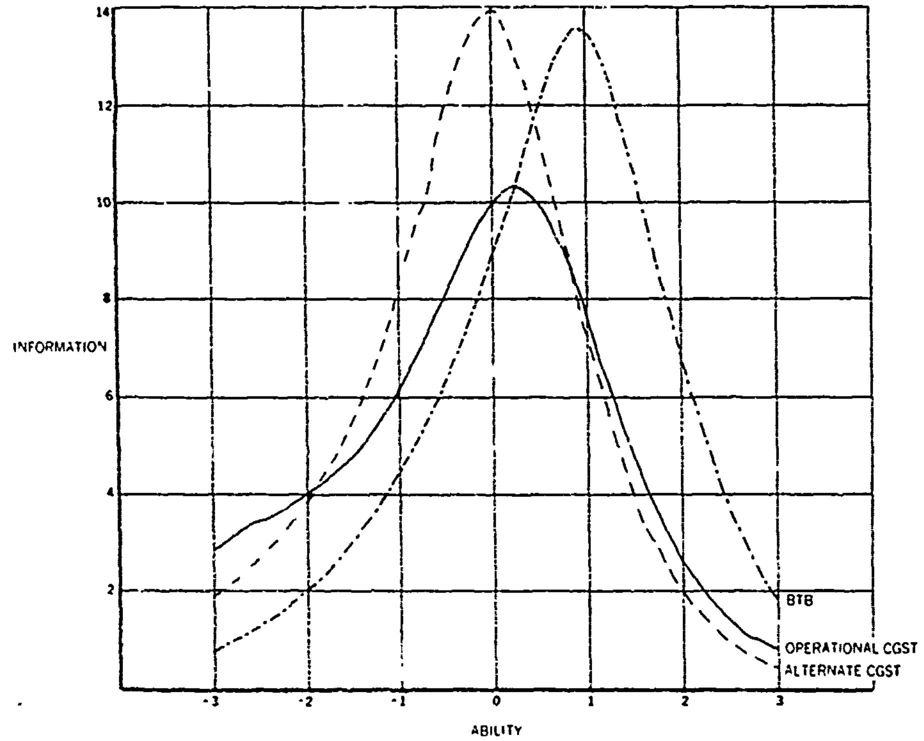


Figure 2. - Test information curves for ARI.

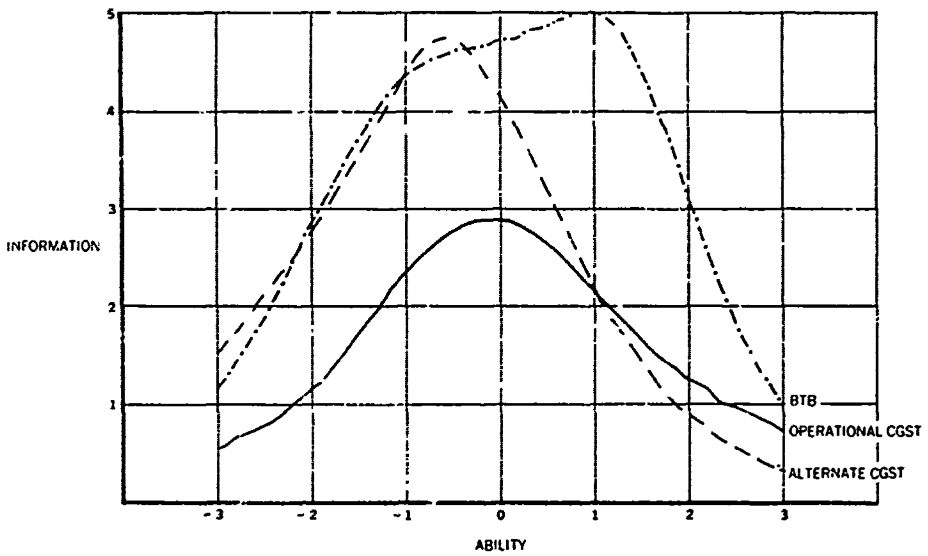


Figure 3.- Test information curves for MEC.

BTB provided more information (i.e., was a more accurate measurement instrument) in most parts of the ability continuum than the operational form of the CGST. This larger quantity of information could be attributable to two factors in test development: number of alternatives per item and differential performance between Blacks and Whites.

The number of alternatives per item is related to the quantity of information per item. A larger number of alternatives decreases the item parameter c which in turn increases the quantity of information provided by the item (Warm, 1978). Since the BTB has five-alternative items as compared to four-alternative items on the CGST for the CCT and ARI, the BTB items should have lower c-values. A comparison of the mean c-values on the BTB and CGST supported this expectation (mean c-values of 0.08 and 0.22 for the GCT on the BTB and operational form of the CGST, respectively; mean c-values of 0.10 and 0.23 for the ARI on the BTB and operational form of the CGST, respectively). The lower c-values on the GCT and ARI on the BTB could partially account for the higher information provided by the BTB's GCT and ARI.

The second factor which might have contributed to the increased information on the BTB is differential performance between groups. When the alternate forms of the CGST were developed, one of the test developers noted that in many instances the items providing the highest information were also among the items identified as potentially biased. Although the interaction between item bias and quantity of information has not been investigated with these data, a comparison of the differential performance between Blacks and Whites, as measured by mean score difference between Blacks and Whites, showed that the differential performance between groups was greater on the BTB than on the operational form of the CGST. On the CGST, the mean scores for Blacks were 14%, 11%, and 18% lower than the mean scores for Whites on the GCT, ARI, and MEC as compared to 25%, 18%, and 20% lower on the GCT, ARI, and MEC on the BTB.[1] Since it is feasible that excluding items identified as potentially biased would reduce the quantity of information on the test, the relationship between item bias and information is worth additional investigation --especially in our current environment which emphasizes test fairness and efficiency in use of resources.

[1]On each of the tests (e.g., GCT on BTB), the difference in performance between Blacks and Whites was significant. Percent differences were used for this type of comparison because of the differences in possible scores on equivalent tests across batteries.

Comparison of Operational and Alternate Forms of the CGST. The tests on the alternate form of the CGST were developed by selecting items which provided high information in the range on the ability continuum near the Coast Guard's cut-score of NSS=45 for each test. Therefore, it was expected that the tests on the alternate form of the CGST would provide more information in this range than the tests on the operational form of the CGST. The data in Figures 1, 2, and 3 confirm that this was the case. On the alternate form of the CGST, the range on the ability continuum equivalent to a NSS range of 44-46 was -1.2 to -0.9 for the VPT, -0.6 to -0.1 for the ARI, and -0.6 to -0.4 for the MEC. As shown in the figures, each test on the alternate form of the CGST provided more information in the selected range than the equivalent test on the operational form of the CGST. This increased information was accompanied by an increase in differential performance between Blacks and Whites. On the operational form of the CGST, mean performance for Blacks on the GCT, ARI, and MEC was 20%, 14%, and 17% lower than mean performance for Whites in the sample of applicants for Coast Guard enlistment. In the same sample, mean performance for Blacks on the VPT, ARI, and MEC on the alternate form of the CGST was 24%, 20%, and 20% lower than mean performance for Whites.

Comparison of the BTB and Alternate Form of the CGST. The differences between the test information curves for the BTB and alternate form of the CGST were less consistent than the differences noted in the previous comparisons. The GCT on the BTB provided more information than the VPT on the alternate form of the CGST. The ARI on the BTB and the ARI on the alternate form of the CGST provided roughly the same quantities of information. The MEC on the BTB provided more information than the MEC on the alternate form of the CGST by providing high information over a wider ability range. In addition to differences in quantity of information, the locus of maximum information on each test varied across batteries.

As with the comparison between the operational and alternate forms of the CGST, the differences in the locus of maximum information can be explained by the method used in developing the alternate form of the CGST. Since the differences in quantity of information between the BTB and operational form of the CGST were attributed to differences in c-values and differential performance between groups, it was logical to assume that the same explanations could be used here. However, comparisons of c-values and differential performance between Blacks and Whites on the BTB and the alternate form of the CGST showed that the differences were quite small. The mean c-values for the GCT, ARI, and MEC on the BTB were 0.08, 0.10, and 0.13 as compared to 0.10, 0.12, and 0.16 on the alternate form of the CGST. The percentage differences in performance between Whites and Blacks for the GCT, ARI, and MEC on the BTB were 25%, 18%, and 20% as compared to 24%, 20%, and 20% on the alternate form of the CGST.

In evaluating the differences between the BTB and the alternate form of the CGST, the following factors should be considered: (1) The examinee data for the two batteries came from two different populations. The data for the alternate form of the CGST came from the Coast Guard's applicant population. The data for the BTB came from a restricted sample -- Coast Guard recruits. (2) The sample size for each test on the alternate form of the CGST was much smaller than the sample size for the BTB -- 700 to 900 for each test on the CGST as compared to approximately 5600 for each test on the BTB. Although the test information curves should not be greatly affected by these differences in samples, the accuracy of the c-values may be slightly affected (Cook and Hambleton, 1979) and the classical statistics are definitely affected.

Summary and Conclusions

The tests on the Navy's BTB provided more information than the GCT, ARI, and MEC on the operational form of the CGST and more information than the VPT and MEC on the alternate form of the CGST. Each test on the alternate form of the CGST provided more information at the Coast Guard's cut-score than the equivalent test on the operational form of the CGST. These differences in quantity of information and locus of maximum information were explained by number of alternatives per item, differential performance between Blacks and Whites, and the method used in developing the alternate form of the CGST. Although the finding that increased information was associated with a larger difference in performance between Blacks and Whites was not consistent across all comparisons, the trend was strong enough to support a recommendation for research into the interaction between quantity of information and item bias.

References

Cook, L. L and Hambleton, R. K. Effects of test length and sample size on the estimates of precision of latent ability scores (Laboratory of Psychometric and Evaluative Research Report No. 87). Amherst, MA: School of Education, University of Massachusetts, 1979.

Hambleton, R. K. and Cook, L. L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14, 75-96.

Marco, G. L. Item characteristic curve solutions to three intractable testing problems. Journal of Educational Measurement, 1977, 14, 139-160.

Rimland, B. The development and standardization of Form 6 of the Navy General Classification Test (Bureau of Naval Personnel Technical Bulletin 58-4). San Diego, CA: U. S. Naval Personnel Research Field Activity, 1958(a).

Rimland, B. The development and standardization of Form 6 of the Navy Arithmetic Test (Bureau of Naval Personnel Technical Bulletin 58-5). San Diego, CA: U. S. Naval Personnel Research Field Activity, 1958(b).

Scheuneman, J. A. A method for assessing bias in test items. Journal of Educational Measurement, 1979, 16, 143-152.

Swanson, L. The development and standardization of Form 6 of the Navy Mechanical Test (Bureau of Naval Personnel Technical Bulletin 58-6). San Diego, CA: U. S. Naval Personnel Research Field Activity, 1958.

Urry, V. W. Ancillary estimators for the item parameters of mental test models. In W. A. Gorham (Chair), Computers and testing: Steps toward the inevitable conquest. Symposium presented at the 83rd Annual Convention of the American Psychological Association, Chicago, 1975. (NTIS No. PB 261 694/AS)

Warm, T. A. A primer of item response theory (Technical Report 941078). Oklahoma City, OK: U. S. Coast Guard Institute, 1978.

AD P000885

Using Generalizability Theory to Assess Interrater Reliability of Contract Proposal Evaluations

Richard A. Kass, Timothy Elig and Karen Mitchell
US Army Research Institute for the Behavioral and Social Sciences

The purpose of this report is to present a technique for estimating interrater reliability in terms of a generalizability coefficient, give an example of this technique from five recent contract proposal evaluations, and present the implications of these data for organizing future contract proposal reviews.

Generalizability Theory

Most investigations of interrater reliability report the product moment correlation between the ratings of the raters. When more than two raters are employed, the product moment correlation may be reported for all possible pairings of raters. There are three general disadvantages with the correlational approach to assess interrater reliability. First, there is a theoretical problem of conceptualizing proposal evaluation scores in terms of the classical notion of true scores. Second, the correlational method does not permit the investigation of different sources of error. Third, when more than two evaluators are involved, pair-wise correlations do not readily allow for estimates of rater reliability based on composite ratings.

Generalizability Theory is an analysis of variance approach to interrater reliability explicated most completely in a book by Cronbach, Gleser, Nanda and Rajaratnam (1972) entitled The Dependability of Behavioral Measurements. Brennan (1977) provides an amplification of the basic principles and procedures.

The first advantage of Generalizability Theory is that it does not rest on the classical notion of true and error scores. Evaluating contract proposals in terms of classical test theory assumes that there is associated with each proposal a true score, and the more (or better) raters employed the better the final observed score will approximate a proposal's true score. In Generalizability Theory, there is no single true score which the evaluators are attempting to approximate. The Generalizability Coefficient (GC) is an index of how well we are measuring (approximating) one particular specified universe out of any number of possible universes of interest.

A universe is a collection of behavioral measurements. A particular set of behavioral measurements in a universe is further defined in terms of the facets or conditions of measurement. With respect to contract proposal evaluations, there are often three facets: raters, criteria and proposals. It will later be shown that the calculation of the GC on the data in this report involves computing a three-factor (facets) completely crossed ANOVA. The "generalizability" (universe of interest) of Generalizability Theory refers to the extent that the facets defining the universe of interest may be fixed or random.

It will be useful to show the relationship between the calculation of the reliability coefficient (Rxx) and the Generalizability Coefficient (GC).

Reliability can be written as:

$$R_{xx} = \frac{\sigma^2(T)}{\sigma^2(T) + \sigma^2(E)} \qquad (1)$$

Where T and E represent true and error scores, respectively.
If we substitute universe score U for true score, the equation for the generalizability coefficient (GC) is:

$$GC = \frac{\sigma^2(U)}{\sigma^2(U) + \sigma^2(E)} \qquad (2)$$

It can be seen that the relationship among the terms remains the same for reliability and generalizability coefficients. The major difference is that the relative size of the U and E terms in the GC formulation will vary depending on the number of facets defining the universe score and whether these facets are considered fixed or random facets.

It was stated earlier that the second major limitation of the correlational approach to interrater reliability is its inability to distinguish different sources of error. In classical test theory there is one complex error term. In Generalizability Theory error variance may be identified for each facet. Estimation of the sources of error variance is most useful in making decisions concerning the design of future contract proposal evaluations. One can answer the question of how much interrater reliability would be affected by increasing or decreasing the number of raters or number of criteria, or both.

The third limitation of the traditional correlational approach is that it becomes awkward when more than two raters are used in the evaluation. The traditional approach is to report the product moment correlation between all possible pairings of raters. In some cases an average or median correlation may be given as a single index for the interrater reliability. There are problems with this approach. An individual correlation between any pair of raters represents the reliability of the evaluation score, if either rater's score was used as the proposal's final score. In practice, this is never done. Both raters' scores are used to yield a composite score. Consequently, the correlation between individual rater's scores is an underestimate of the reliability of the composite score. Since all correlations between possible pairs of ratings are underestimates, the average or median of these correlations will be an underestimate also. The extent to which the correlation underestimates the reliability of a composite score increases as the number of raters increases. The Generalizability coefficient provides an index of the reliability of the composite rating. In this manner it may be noted that generalizability coefficients are interclass correlations (Ebel, 1951). Generalizability Theory, however, is an expansion of the interclass coefficient approach to allow for more complex experimental designs.

TABLE 1

A Rater by Criterion by Proposal (PxRxC) ANOVA Design

RATERS

CONTRACT PROPOSALS	R1					R2					R3					R4				
	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
P1																				
P2																				
P3																				
P4																				
P5																				
P6																				
P7																				
P8																				
P9																				

TABLE 2

ANOVA Summary Table for the (PxRxC) Design

Source	df	SS	MS
Proposals (P)	8	791	98.9
Raters (R)	3	121	40.3
Criterion (C)	4	8,809	2,202.2
PR	24	357	14.8
PC	32	1,015	31.7
RC	12	876	73.0
PRC	96	545	5.7

An Empirical Example

In this section, the interrater reliability of five different sets of contract proposals are analyzed using the generalizability theory approach. The contract evaluations are actual evaluations conducted at the US Army Research Institute (ARI) and they vary along the following dimensions:

Contract Proposal Evaluation Set	Number of Proposals Evaluated	Number of ARI Raters	Number of Criteria Used
A	3	5	3
B	6	3	3
C	8	4	4
D	9	4	5
E	31	3	4

To illustrate the ANOVA method, the interrater reliability of contract proposal evaluations set "D" is worked out in a step-by-step fashion. Table 1 depicts set "D" contract proposals evaluation in terms of a three-way ANOVA experimental design. Nine proposals were received, four raters were used. Each rater (R) rated all proposals (P) with respect to five criteria (C). These criteria reflect separate ratings for different aspects of the proposals, for example, technical adequacy, organizational experience, etc. Accordingly, each proposal received a total of 20 ratings (4 raters x 5 criteria).

In contract proposal evaluations, raters are considered a random facet so that the final evaluation scores will generalize to the use of other raters having similar levels of expertise. The criterion facet is considered a fixed facet in that the final evaluation scores do not generalize to other criteria. That is, the use of some other criteria for a proposal evaluation may result in a different final rank ordering of the proposals.
The proposals facet is considered a random facet in that having more or fewer proposals would not change the score assigned to any one proposal.

Table 2 presents the traditional ANOVA summary data for the actual ratings obtained in the proposal evaluation. In the traditional ANOVA, emphasis is on the statistical tests of the "main" and "interaction" effects by selecting the ratio of the appropriate Mean Square effect and appropriate Mean Square error term. In Generalizability Theory the ANOVA summary table is used only to obtain the quantities for the Mean Squares.

The next step is to compute the unique variance estimates for each facet using data in the ANOVA summary table and the formulations of the components of the Expected Mean Squares. Fortunately, there are well worked out procedures for this (Brennan, 1977). The final variance estimates for the separate facets are presented in Table 3 under the column for G-study variance estimates.

Generalizability theory distinguishes between G studies and D studies. G studies are oriented towards obtaining estimates of the various sources of error variances and G studies are characterized by random-effects ANOVA models. D studies, on the other hand, are designed to determine variance estimates in an

TABLE 3

Changes in Interrater Reliability Due to Changes in the Number or Raters or Criteria

Component	G Study Variance Estimate	Possible D Studies										
		R=4 C=5		R=6 C=5		R=3 C=5		R=4 C=3		R=4 C=7		
		N	σ^2	N	σ^2	N	σ^2	N	σ^2	N	σ^2	
Proposals (P)	2.9	1	2.90	1	2.90	1	2.90	1	2.90	1	2.90	U
Raters (R)	0											
Dimensions (C)	58.3											
PR	1.8	4	.45	6	.30	3	.60	4	.45	4	.45	E
PC	6.5	5	1.30	5	1.30	5	1.30	3	2.17	7	.93	U
RC	7.5											
PRC	5.7	20	.29	30	.19	15	.38	12	.48	28	.20	E
Total Universe Variance (U)			4.20	-	4.20		4.20		5.07		3.83	
Total Error Variance (E)			.74	-	.49		.98		.93		.65	
Generalizability Coeff (GC)			.85	-	.90		.81		.85		.85	

TABLE 4

Computed Generalizability Coefficient
for Each of the Five Contract Proposal Evaluations

Number of Raters

NUMBER OF CRITERIA	3	4	5
3	Data Set B .99		Data Set A .95
4	Data Set E .88	Data Set C .94	
5		Data Set D .85	

actual situation where some facets of the ANOVA model are fixed. While our empirical example is a D study, the results can be used to estimate G-study variances by temporarily assuming that the three facets are random effects. These estimated G-study variances can, in turn, be used to estimate variances for various D-study configurations of interest. The individual D study variance estimates are obtained by dividing the G-study variance estimates by their respective sampling frequencies. The D-study universe (U) and error (E) variances are combined according to equation 2 to compute the GC. For data set "D" with four raters (R = 4) and five criteria (C = 4), the generalizability coefficient is .85.

<u>Extrapolation of Data Set "D" to Other Evaluation Designs</u>

One can compute the extent of expected change in the GC when either (or both) the number of raters or number of criteria is changed. The necessary computation is quite easy. To determine the effect on interrater reliability of increasing the number of raters from four to six, the sampling frequency (N) is changed accordingly and the G-study variances are divided by the new sampling frequencies. This procedure is equivalent to using the Spearman-Brown prophesy formula to determine increases in reliability as test length is increased.

Data in Table 3 summarize changes in the GC for data set "D" when the number of raters or criteria is changed. Increasing or decreasing the number of raters directly increases or decreases the GC. This is because both ANOVA components involving raters contribute to the error term. This may be contrasted to the negligable effect resulting from changes in the number of criteria. Since criteria contribute to both the universe score variance and error variance, the GC ratio of these two terms changes little.

<u>Extrapolation to Other Evaluation Designs Using All Five Data Sets</u>

The projected changes in interrater reliability in Table 3 are based on the G-study variance estimates from one data set. Estimates of the effects of increasing and decreasing the number of raters and/or criteria on interrater reliability are strengthened to the extent that more G-study variance estimates are obtained. The procedure outlined for data set "D" was applied to the other four data sets. The computed generalizability coefficients for all five data sets are presented in Table 4.

The information in Table 4 can be used to compute the effects on reliability of changing the number of raters and/or criteria. Five replications of Table 4 can be estimated by using each data set independently to estimate changes in GC due to changes in the number of raters and criteria. Combining these five sets of independent estimates would yield five interrater reliability coefficients in each cell of Table 4. Moreover, the table can be expanded to provide estimates for combinations of one to seven raters and one to seven criteria. For comparison purposes, the means for each cell have been plotted in Figure 1.

Figure 1 indicates that as the number of raters used in the evaluation increases, so does the interrater reliability. The rate of increase decreases, however, as the number of raters exceeds five. In a similar manner, there is little effect of increasing the number of criteria beyond three. These data suggest for similar evaluations an average level of interrater reliability of .90 can be attained by using three raters and three criteria per contract proposal evaluation.

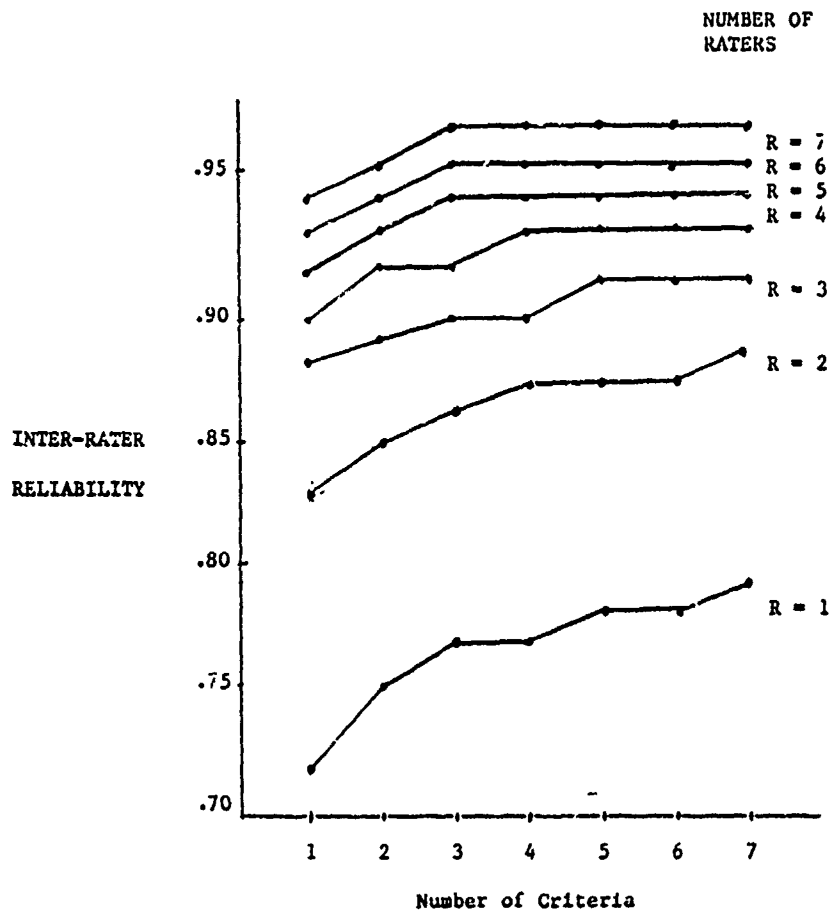


FIGURE 1. Interrater Reliability as a Function of the Number of Raters and Criteria.

REFERENCES

Brennan, R. L. Generalizability analyses: principles and procedures (ACT Technical Bulletin No 26). Iowa City, Iowa: The American College Testing Program, 1977.

Cronbach, L. J. K., Gleser, G. C., Nanda, H.A.N. and Rajaratman, N. The dependability of behavioral measurements. New York: John Wiley and Sons, Inc., 1972.

Ebel, R. L. Estimation of reliability of rating. Psychometrica, 1951, 16, pp. 407-424.

Using Automated Personality Test Interpretation for Security Screening

Benjamin Kleinmuntz
University of Illinois at Chicago

Introduction

Automated decision rules for personality test interpretation have been in use since the mid 1960's. About a dozen such systems are currently available, mainly for psychodiagnosis and to plan treatment for psychiatric patients in hospitals and clinics. One of the first such automated decision rule systems was developed at Carnegie-Mellon University (Kleinmuntz, 1963). It demonstrated that a computer can be programmed to simulate and even outperform the human clinician whose interpretive strategies it originally modeled.

The system so developed was based on the "thinking aloud" protocol of an expert interpreter sorting through 126 Minnesota Multiphasic Personality Inventory (MMPI) profiles of emotionally maladjusted and adjusted college students. The protocol of an expert MMPI interpreter, plus information subsequently drawn from the MMPI research literature, served as a data-base for developing the decision rules. The final set of interpretive rules also outperformed many other clinicians in a crossvalidation study that pitted the computer's ability against that of human expert interpreters (Kleinmuntz, 1969).

The purpose of this paper is to describe a similar MMPI automated interpretive system that was specifically designed for a nonpsychiatric population, particularly for purposes of screening personnel who will hold sensitive positions in paramilitary and military settings. But first a brief description of the process of automating human judgments.

Background

The idea of borrowing, simulating, or modeling the decision strategies of experts in particular specialties has its origins in the work of the information processing group at Carnegie-Mellon University (see Newell & Simon, 1972). Automating intelligent behavior has been applied in a variety of areas including chess (de Groot, 1965), symbolic logic (Newell & Simon, 1961), cryptarithmetic (Newell & Simon, 1972), physics (Langley, 1979; Simon & Lea, 1979) and medicine (Kleinmuntz, 1983). Essentially, the task of the researcher in computer simulation studies is to attempt to learn from an expert diagnostician or decision maker how that person makes decisions - i.e., what strategies and information the expert uses to arrive at problem solutions. Thus, if we were interested in simulating a military intelligence officer's decision strategies, we might design a task environment that encourages him to think through (and aloud) his problem from its beginning to its solution. If the officer is a good problem solver as judged by some predetermined criterion of a correct solution, then we have a tape or video recorded trace of the correct solution path; likewise, if the officer is not a good problem solver, we obtain a trace of the incorrect path. Clearly, this permits modeling, automating, and comparing good and poor decision makers.

Method

The information processing or simulation approach is a lengthy and some-

times unfeasible procedure. It requires many hours of laboratory work and, although it may yield important fine-grain data, it may not be practical for developing decision rules that need to be designed for immediate application. Therefore, in our study, because there was a need to develop a set of decision rules quickly for an existing population, we used a more direct method of obtaining a data-base of decision strategies. Our method relied heavily on existing research literature for information about how clinicians arrive at MMPI interpretive decisions.

Consequently, the MMPI research literature was examined in order to locate as many descriptive adjectives as possible that are commonly associated with particular profile elevations. This search uncovered 500 such descriptive statements, which were then narrowed down to 162 descriptions that were deemed appropriate to settings that employed personnel for sensitive assignments.

In order to determine empirically whether these descriptions in fact would be helpful for personnel screening in the particular settings where they might be used, a questionnnaire was sent to administrators of police officers (state, municipal, and county), security guards, correctional officers, nuclear plant employees, and military intelligence personnel. At each of these 331 installations, clinical psychologists and other supervisory administrators rated each of the 162 adjectives as either "Useful", "Somewhat Useful", or "Useless."

Results

An analysis of these questionnaire ratings disclosed that 88 descriptive statements were considered very useful by a large majority of respondents. The adjectives were then divided into the following categories, according to their relevance: 1) Test Taking Attitudes (e.g., "is dishonest about self-description"); 2) Attitudes Toward Others (e.g., "is devious in dealing with people); 3) Work Attitudes (e.g., "is not alert, capable, and responsible"); 4) Emotional Factors (e.g., "can be sadistic"); 57 Decisiveness (e.g., "has difficulty making decisions"); 6) Areas Requiring Further Investigation (e.g., "may use hard drugs excessively"); and 7) Overall Evaluation (e.g., "is not trustworthy", "do not hire", "needs psychological counseling").

In the interpretive decision rules that emerged, each of the adjectives is associated with its appropriate decision rule and is printed out by the computer if and only if an individual's MMPI profile has particular clinical scale elevations. And depending on the extent of the elevations — expressed in terms of T-scores that have means of 50 and standard deviations of 10 — each of the relevant descriptive statements is assigned a 5-point rating that reflects the height of the respondent's MMPI scale elevations. These ratings range from a relatively low (T-score of 70) to very high (T-score of 90) levels and are calibrated along a continuum of half standard deviation increments beginning from T-score of 70 (e.g., T-scores of 70, 75, 80, 85 and 90). Hence, the least severe manifestation of a behavior reflects an appropriate elevation(s) of a T-score of 70, whereas the most severe form reflects an elevation(s) of a T-score of 90.

Thus, an automated MMPI report for an individual consists of a printout of that person's MMPI raw and T-scores, a profile plot, scores on certain special research scales, an interpretive and calibrated description of the

meaning of that person's scores, a narrative that reflects the respondent's possible need for psychological counseling, and an evaluation of assessee's probable risk for assignment to a sensitive position.

Discussion

The MMPI interpretive decision rules, which were developed under contract with London House Management Consultants, Inc. of Chicago, are now being normed, validated, and crossvalidated in several studies being conducted in a variety of settings. These studies are taking place among Washington and Wisconsin State Police, Honolulu Police Department, a number of security guard firms, and the United States Army Military Intelligence Command at Fort George G. Meade, Maryland. Consequently, the automated systems are being tailored to the specific needs and cut-off scores prevailing in these settings. Future studies will be aimed at collecting similar data in other settings.

The really unique feature of the London House automated MMPI decision rule system is that it is based on the surveyed needs of users and therefore does not rely merely on the testimonials of satisfied customers. As such the present system was devised according to one of the pivotal guidelines of the Equal Employment Opportunity Commission, which has over the years stipulated that tests and interpretations of test scores take into consideration the demand characteristics of the employment environment. In this case, since the MMPI decision rules were designed according to Employers' specifications of undesirable and potentially damaging personality and attitudinal characteristics, the decision rules are content validated in terms of these employment demands. The empirical validation of the rules, as indicated earlier, is now being conducted.

Summary

Automated MMPI decision rules for the detection of prespecified undesirable personality attributes in sensitive assignments were developed. The settings included police, security, correctional, and military intelligence installations where it is important to identify potential disruptive personalities. The system consists of 88 descriptive statements plus several other evaluative interpretations that are quantitatively assigned to MMPI scale elevations. The system was developed in the context of an information processing approach to computer thinking.

References

de Groot, A.J. Perception and memory versus thought: Some old ideas and recent findings. In Kleinmuntz, B. (Ed.). Problem solving: Research, method and theory. N.Y.: Wiley, 1966.

Kleinmuntz, B. Personality test interpretation by digital computer. Science, 1963, 139, 416-418.

Kleinmuntz, B. Personality test interpretation by computer and clinician. In Butcher, J.N. (Ed.). MMPI: Research developments and clinical applications. N.Y.: McGraw-Hill.

Kleinmuntz, B. Diagnostic problem solving: A historical review and the current state of the science. Medical Decision Making, 1983, 3 (In press).

Langley, P. Rediscovering physics with BACON-3. Proceedings of the 6th IJCAI-Tokyo, 1979, 1, 505-507.

Newell, A. & Simon, H.A. Computer simulation of human thinking. Science, 1961, 134, 2011-2017.

Newell, A. & Simon, H.A. Human problem solving. N.J.: Prentice-Hall, 1972.

Simon, H.A. & Lea, G. Problem solving and rule induction. In Gregg, L.W. (Ed.). Knowledge and cognition. N.J.: LEA, 1974.

AD P000887

Importance of Personal Characteristics to Job Performance

William R. Koch

The University of Texas at Austin

Introduction

As is so often the case, practical considerations served as the primary motivation for the present research. Administrative concern had surfaced over the need to update the existing methods used to screen and select applicants for the position of Guard for the University of Texas Police Department (UTPD). The desire was to develop entry-level assessment procedures that would assist in predicting subsequent job performance levels and in reducing high employee turnover rates, while still complying with the Uniform Guidelines on Employee Selection Procedures (1978).

Efforts began at the ground level because the job description for the position of Guard was quite general, and no task analysis of the position had been performed. Furthermore, literature on job descriptions and tasks of city and state police officers could not be generalized to the university setting due to the quite different functions of campus peace officers. Also, the position of Guard was further differentiated from that of Commissioned Officer within the UTPD. Guards were expected only to have the equivalent of a high school diploma and were responsible for parking enforcement, building security, and public relations posts; Commissioned Officers were expected to have a minimum of two years of college, were responsible for law enforcement and crime prevention, and were authorized to carry weapons.

Preliminary Procedures

It was clear that a detailed job content analysis would be required, not only to specify the requisite knowledge, skills, abilities, and personal characteristics for successful performance of the job tasks, but also to develop a list of criteria to indicate job performance levels for use subsequently in predictive validity studies. An extensive program of structured personal interviews was conducted with a wide cross-section of employees and supervisors. The purpose was to obtain information as to the specific tasks performed by the Guards, the relative amounts of time spent on the various tasks, and the knowledge, skills, and personal abilities necessary to perform the tasks. Detailed notes were taken during the interviews and later were used to construct a questionnaire consisting of items that reflected traits of the employees which were consistently mentioned as being important to job performance.

Upon completion of pilot testing and item revisions, the final form of the questionnaire consisted of 32 items. The same seven-point Likert response format was chosen for use with each item on the questionnaire. Respondents were asked to rate each item in terms of its perceived importance to successful job performance, ranging from "Not Important" at one end of the scale to "Extremely Important" at the other end.

The basic purpose of administering the questionnaire was to attempt to determine, from the viewpoint of the employees themselves, the general dimensions or factors underlying successful job performance. It was felt that the identification of such dimensions would facilitate the subsequent selection (or development if necessary) of an appropriate instrument that could eventually be tried out and possibly validated for entry-level employee selection purposes. For example, one instrument widely used as part of an employee assessment battery has been the Survey of Interpersonal Values (Gordon, 1976), intended to measure the values that people deem important in their interpersonal relationships in terms of six separate factors.

Factor analysis has been the procedure used traditionally to attempt to determine the dimensionality or number of factors being measured by the items on a Likert-type questionnaire. However, factor analytic procedures make some fairly restrictive assumptions about the input data, assumptions that may not be met in practice. For example, factor analysis assumes that the input variables have continuous rather than discrete distributions, that the variables are measured on so-called equal interval scales, and that the relationships among the variables are linear. Moreover, the often used rule of thumb is that the ratio of observations to variables should be at least ten to one for stable results.

When the input data fail to meet the conditions for factor analysis, as they certainly do in the present research, a viable alternative for analysis may be nonmetric multidimensional scaling (MDS). No distributional assumptions about the variables need to be made, the variables need only be measured on ordinal scales, and the relationships among the variables are merely assumed to be monotonic.

In MDS the input data frequently take the form of proximity ratings (Shepard, 1972) which indicate the similarities, dissimilarities, or other associations among the variables. The underlying goal of the procedure is to approximate a one-to-one relationship between the ordinal information in the input data and the corresponding rankings of the distances among the variables represented as points in space. Also, similar to factor analysis, dimensions of the space may sometimes emerge that represent the variations among individuals' ratings of stimuli. However, unlike factor analysis, MDS is not restricted to a vector or axial representation of the space. Frequently, regional interpretations of the spatial configuration are the most obvious and useful (Levy, 1981).

Method

The questionnaire was administered by mail to all employees in the Guard position, with instructions for them to complete the form individually and return it directly to the present author. Anonymity of responses was assured. A sample of 53 out of 57 total Guards returned the survey, a very good response rate.

Prior to the MDS analysis the 53 respondents to the questionnaire were divided into two separate groups consisting of those who were relatively new employees (less than two years on the job) and those who were longer

term employees (more than two years). It was hypothesized that systematic differences might exist in the relative importance placed by each of these two groups on the various items on the questionnaire.

In order to obtain dissimilarity data to serve as input to the MDS run, the rectangular array of raw data was transformed using the formula

$$d_{ij} = \frac{1}{N} \sum_{a=1}^{N} (x_{ia} - x_{ja})^2 \quad \text{for } i,j = 1,\ldots,n \text{ items}$$

where d_{ij} was the dissimilarity between the scores (ratings) for item i and item j across the N persons in each group. This transformation created two square, symmetric matrices of dissimilarities based on 29 persons in group 1 and 24 persons in group 2. These matrices were input to the nonmetric individual differences MDS option of the ALSCAL-4 computer program (Young & Lewyckyj, 1979).

Results

The program parameters were set to request that solutions be performed in from two to six dimensions so that the proportion of variance accounted for could be compared across solutions. The stress (Kruskal, 1964) for each solution was also obtained, although the stress coefficient is not strictly applicable to individual differences scaling problems. It is common practice in MDS analysis to use parsimony, interpretability, and visualizability as the primary criteria in deciding on the dimensionality of the solution. Taking all of these criteria into account, the three dimensional solution, accounting for 85% of the variance in the data (and stress = .21), was deemed to be optimal.

Table 1 below contains the coordinates for each item, relative to three orthogonal reference axes, that locate the items in the spatial configuration. The actual text of each item is also presented. The reference axes in MDS solutions are purely arbitrary and usually do not correspond to the actual dimensions underlying the relationships among the items (stimuli). Rather, the patterns of the positions of the points in space must be examined in detail in an effort to detect meaningful groupings of the items into regions or general dimensions of item variation. This is especially necessary in the absence of outside criteria that might otherwise be regressed on the coordinates to facilitate the interpretation of the configuration and the labeling of dimensions (Kruskal & Wish, 1978).

A subjective process of classification of variables can often be quite helpful as a preliminary step in attempting to interpret the MDS solution. Hypothetical dimensions along which the stimuli vary are formulated and then tested by examining the projections of the stimuli onto various directions through the space. It is important to consider the ordering of the stimuli along dimensions, the distances between the stimuli, and the stimuli that are positioned at opposite ends of dimensions. All of these procedures were used in interpreting the MDS configuration in the present research.

Table 1

Coordinates for the Questionnaire Items

Item	Reference Axis 1	Reference Axis 2	Reference Axis 3
1. leadership ability	1.28	1.82	3.20
2. being dependable	-1.12	-1.23	- .77
3. job knowledge	- .18	-1.05	-1.09
4. maturity in behavior	- .16	-1.08	- .32
5. educational background	3.01	2.89	.89
6. ability to accept criticism	1.03	.53	-1.17
7. tolerance in dealing with public	-1.12	- .85	-1.10
8. coping with stress	1.32	.61	- .15
9. common sense	-1.25	- .87	-1.29
10. asserting yourself	.32	1.38	.15
11. taking pride in your work	- .68	.05	1.03
12. being observant	- .25	.09	- .26
13. punctuality	- .78	- .79	- .96
14. helping others	.06	- .41	.41
15. ability to work alone	.06	- .96	.74
16. setting a good example	- .86	.61	.44
17. quick thinking	- .23	- .15	.54
18. reliable job attendance	-1.11	- .54	- .31
19. being easy going or relaxed	.41	1.57	1.08
20. courtesy toward others	-1.05	- .38	- .59
21. alertness	- .12	- .31	.23
22. good judgment	- .87	- .46	- .23
23. honesty	-1.25	- .95	- .50
24. controlling your temper	- .99	- .76	- .63
25. verbal or communication skills	.59	- .15	- .15
26. personal appearance	.05	- .64	.26
27. sensitivity to others' feelings	.00	.65	.62
28. memory for names, faces, numbers	2.01	.65	- .59
29. firmness in dealing with others	- .46	1.44	- .67
30. familiarity with the campus	.52	-1.26	- .94
31. tolerance for boredom	.79	- .29	2.62
32. following written procedures	1.02	.86	- .49

Three distinctive regions or clusterings of items were discovered in the space, each having the general shape of an ellipsoid. The major axis of the first ellipsoid defined an interpersonal relations dimension, the major axis of the second ellipsoid defined a cognitive skills dimension, and the major axis of the third ellipsoid defined a work ethic dimension. All three of the dimensions were bi-polar in the sense that the endpoint items were somewhat opposite to each other in meaning. However, these dimensions were not orthogonal. Rather, the interpersonal relations dimension was highly related to the cognitive skills dimension and moderately related to the work ethic dimension. Also, the cognitive skills dimension was slightly related

to the work ethic dimension. The angles formed by the major axes to each other were used to estimate these interrelationships. Table 2 below illustrates the items classified into each of the three ellipsoid regions and the item locations along the major axis of each.

Table 2

Classification and Locations of Items

Location	DIM 1 Interpersonal Relations	DIM 2 Cognitive Skills	DIM 3 Work Ethic
High Positive	1. leadership ability 19. being easy going or relaxed 10. asserting yourself 29. firmness in dealing with others	5. educ. background 28. memory for names, faces, numbers 32. follow written procedures	31. tolerance for boredom
Neutral	8. coping with stress 6. ability to accept criticism 27. sensitivity to others' feelings 14. helping others	25. verbal or communication skills 17. quick thinking 21. alertness 12. being observant	11. pride in work 15. ability to work alone 16. setting a good example 26. pers. appearance
High Negative	20. courtesy toward others 24. controlling your temper 7. tolerance dealing with public	30. familiarity with the campus 3. job knowledge 22. good judgment 9. common sense	4. maturity in behavior 18. reliable job attendance 23. honesty 2. dependability 13. punctuality

The positive end of the interpersonal relations dimension reflected assertive, self confidence behaviors, while the negative end reflected self control behaviors, with all of the items related to interactions among persons. The positive end of the cognitive skills dimension denoted formally learned skills, whereas the negative end denoted on-the-job learning and basic common sense, and all of the items can be classified as cognitive skills. The positive end of the work ethic dimension was defined by ability to tolerate boredom on the job, while the negative end was defined by traditional work ethic behaviors like honesty, dependability, punctuality, etc. Again, all of the items can be classified intuitively as manifestations of work ethic behaviors.

The subject weights derived from the individual differences scaling (see Schiffman, Reynolds, & Young, 1981), indicating the relative importance placed on the dimensions by inexperienced compared to experienced employees, revealed

differences only with respect to the third dimension. Inexperienced Guards placed substantial importance on the work ethic dimension, while the weight for the experienced Guards was close to zero. It was not surprising to find that new workers, eager to make a good impression, were quite concerned with the importance of such traditional indicators of job performance as being dependable, punctuality, and reliable job attendance.

Conclusion

The results of the present research have considerable utility in and of themselves. Namely, in developing an instrument or set of rating criteria for use by supervisors to evaluate employee job performance, it is valuable to know the three basic dimensions that the employees themselves believe to be important to doing a good job. Procedures for employee selection for the position of Guard would also need to be focused on the assessment of interpersonal relations abilities, cognitive skills, and work ethic behaviors. However, beyond the practical considerations of the present research, the advantages of MDS methods were demonstrated in their ability to discover underlying relationships among variables while making only minimal assumptions. MDS methods will likely find many applications to a wide variety of measurement requirements arising in the context of personnel assessment.

REFERENCES

Gordon, L.V. Survey of interpersonal values. Chicago: Science Research Associates, Inc., 1976.

Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27.

Kruskal, J.B. & Wish, M. Multidimensional scaling. Beverly Hills: Sage Publications, 1978.

Levy, S. Lawful roles of facets in social theories. In I. Borg (Ed.), Multidimensional data representations. Ann Arbor: Mathesis Press, 1981.

Schiffman, S.S., Reynolds, M.L., & Young, F.W. Introduction to multidimensional scaling. New York: Academic Press, 1981.

Shepard, R.N. A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R.N. Shepard, A.K. Romney, & S.B. Nerlove (Eds.), Multidimensional scaling: theory and applications in the behavioral sciences (Vol. I). New York: Seminar Press, 1972.

Uniform guidelines on employee selection procedures. Federal Register, 1978, 43, 38290-38315.

Young, F.W., & Lewyckyj, R. ALSCAL-4: User's guide. Carrboro, NC: Data Analysis and Theory Associates, 1979.

Enlisted Military Selection: Impacts of Changing Aptitude Standards Since 1940 [1,2]

Janice H. Laurence, Brian K. Waters, and Linda S. Perelman
Human Resources Research Organization

Uncle Sam wanted "you" in 1940 if you had the ability to comprehend simple orders given in the English language. Today, considerably greater evidence of training aptitude is required of military applicants. Mental standards for entry into the Military Services have become more stringent, or at least more sophisticated, over the past four decades.

Since World War II, military technology (e.g., weapons systems) has become increasingly complex and, as a result, greater mental and educational demands have been placed upon enlisted personnel. The Navy, for example, cannot rely solely on brawny seamen to fulfill its mission. Rather, it needs technical specialists to man nuclear powered ships, to maintain aircraft, and to operate radar devices. The demand for personnel quality, above and beyond basic literacy, has prompted the Services to employ more complex psychometric screening and classification devices to determine which individuals have the capacity for efficiently absorbing training and becoming effective soldiers, sailors, marines, or airmen.

Currently, the Army will let you "be all you can be", the Navy will let you "see the world", the Marine Corps will consider you to be one of the "few good men", and the Air Force will let you "fly with them" if you meet their particular mental requirements based upon aptitude test scores in conjunction with educational status. Each Service designates minimum acceptable Armed Forces Qualification Test (AFQT) scores and (with the exception of the Navy) specific Armed Services Vocational Aptitude Battery (ASVAB) aptitude area scores separately for high school graduates and non-graduates wishing to enlist.

Although aptitude standards for entry into the Armed Services are much higher today than they were before World War II, they have not increased monotonically since that time. Selection criteria for induction and enlistment into the military have been adjusted many times since 1940 in response to a number of factors, in addition to the military's technological demands. Factors both internal and external to the military (e.g., manpower requirements and national economic conditions) have at times necessitated temporary "lowerings" or enabled the raising of aptitude requirements for military service.

The present report historically tracks the changes in minimum aptitude qualifications for military service and discusses some of the factors accompanying such changes. Standards have changed in the past and they are likely to change in the future. An historical track of the shifts in minimum qualifications may enable manpower analysts to recognize the conditions which may lead to lower or more complex aptitude standards.

Definitions of Standards and Quality

Selection standards are the criteria below which individuals may not be accepted for induction or enlistment into a Military Service. The basic purpose of such standards is to screen out potential enlisted personnel who are least likely to profit from training and who might be actual liabilities to the Services.

Beginning in 1946, entry aptitude standards were expressed in terms of minimum scores on standardized tests in addition to the previous literacy requirements. Since the mid 1960s, standards have differed according to educational attainment. That is, minimum qualifying scores are used in conjunction with educational level to determine whether an examinee is eligible to serve in the Armed Forces. Today, for example, non-high school graduates and GED recipients are required to achieve higher scores on the AFQT than high school graduates to be considered for military duty.

Under varying DoD limitations, the individual Services, due to their unique missions, technical requirements, and recruiting market experience, set the standards below which individuals are not eligible to enlist. Meeting Service minimum standards, however, does not guarantee entry into the military. From time to time, the Services set higher quality goals and temporarily adjust applicant qualification requirements through more selective operational "cutting scores." These are a less definable set of decision rules which operate on a daily basis to regulate the flow of lower quality personnel.

The Services prefer "high quality" personnel. They seek to recruit and select as many high school graduates and persons scoring at or above average on the AFQT as manpower requirements demand and the labor market supplies. When there is a reduction in numerical requirements and/or when the recruiting market shows ample supply of top quality applicants, these higher "cutting scores" operate to select the best from the applicant pool. While lower quality personnel do enter the system, their numbers are greatly reduced. As is common in civilian hiring practices, military recruitment procedures move toward groups previously excluded or numerically limited (by policy) in a tightening market and either qualify individuals nearer the existing minimum standards or adjust the standards downward under extreme conditions (e.g., war).

---

[1]This paper is an abbreviated version of a forthcoming technical report for the Office of Naval Research.

[2]Paper presented at the 24th Annual Meeting of the Military Testing Association, San Antonio, TX: November 1982.

The recruiting market of the past fiscal year (1982) was one in which all four Services could afford to be choosy in selecting recruits. Informal enlistment standards operated at a relatively high level and good quality among accessions resulted. A recruiting boom such as this will not last forever. In the past, quality has often been the first to suffer in an unfavorable selection environment. Perhaps it is possible to learn our lessons from the past and prepare for a decline in the number of military applicants without incurring the risks involved in an extreme reduction in the proportion of well qualified personnel.

Simplifying Complexity: A Model of Factors Influencing the Selection Process

Service enlistment policies and hence the quality of military accessions depend upon the interplay of environmental factors, both internal and external to the military. Figure 1 shows some of the many factors which influence Service minimum and operational selection standards and the quality mix of accessions.

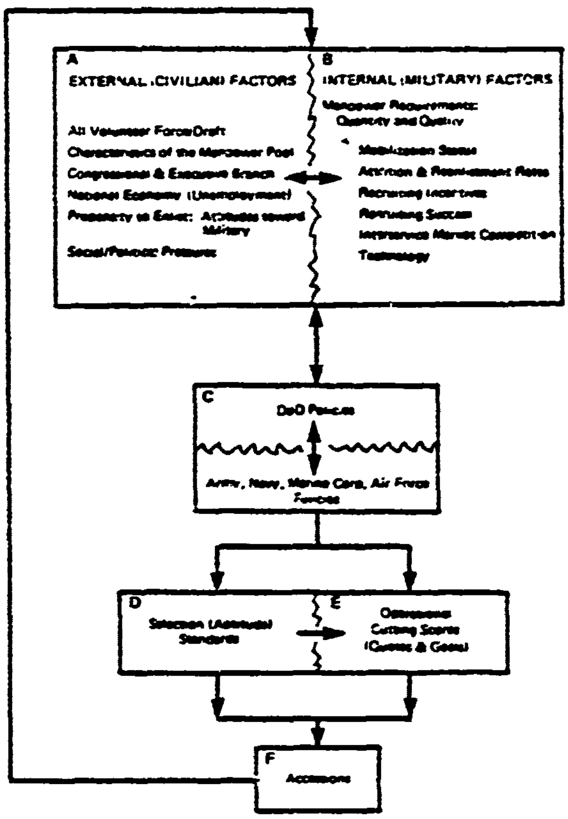


Figure 1. HumRRO Model of Factors Influencing Selection Aptitude Standards and Characteristics of Accessions

The military selection process, while at all times trying to maximize quality, operates within the context of external (i.e., civilian) constraints, examples of which appear in Box A of Figure 1. These factors are briefly delineated below.

- All-Volunteer Force/Draft - National policy on the establishment of an AVF as opposed to compulsory service has the greatest effect, of any single factor, on the quality of examinees and required recruiting resources to meet strength objectives.
- Characteristics of the Manpower Pool - The military draws its recruits primarily from civilian male youth ages 18 to 23. The number and aptitude levels of such youth, for example, are major recruiting market considerations.
- Congressional & Executive Branch Activities - Congress and/or the executive branch may place legal and/or policy constraints on military service selection.
- Defense Budget Appropriations - The level of funding and programmatic decisions in the defense budget process directly affect manpower programs.
- National Economy (Unemployment) - There is direct correspondence between the youth unemployment rate and the quantity of military applicants. Operational cutting scores can be adjusted so as to produce a large proportion of top quality military accessions.
- Propensity to Enlist: Attitudes Toward Military - Favorable attitudes toward the military in general and towards specific Services greatly affect the likelihood that an individual will enlist.

- Social/Political Pressures - Generally, equal opportunity considerations come into play here (e.g., utilization of women and minority representation). An example is pressure to involve the military institution in social rehabilitation for the underskilled and undereducated. Standards and/or cutting scores may be adjusted downward to accommodate such pressures.

These external factors effect and in turn are affected by factors within the military. As depicted in Box B, and elaborated on below, these internal factors generally are related to or subsumed under military manpower requirements.

- Mobilization Status - Force strength objectives are primarily driven by war/peace preparations. During wartime mobilizations, for example, standards may be lowered to qualify more men in the face of drains on available manpower.
- Attrition & Reenlistment Rates - The number and type of recruits needed tomorrow are direct functions of the retention behaviors of the enlisted personnel of today.
- Recruiting Incentives - Enlistment bonuses, educational benefits, and assignment options can affect the attractiveness of a military Service to potential recruits.
- Recruiting Success - Tomorrow's recruiting goals are an inverse function of today's recruiting outcomes.
- Interservice Market Competition - The relative attractiveness of one Service to potential recruits impacts upon the quality of personnel available to the other Services. For example, the perceived desirability of the Air Force negatively impacts the number of high quality Army applicants.
- Technology - As military weapons systems become more complex the need for well qualified recruits to operate them increases.

Although these factors have been discussed separately, they interact to effect DoD and individual Service policies (Box C) in setting selection aptitude standards and operational cutting scores (Boxes D & E) which, in turn, determine the quantity and quality of military accessions (Box F). Finally, as shown by the feedback loop from Box F to A, the accessions which result from the complex selection process have an impact on the external and internal driving factors. For example, the high levels of youth unemployment in FY 1982 are assumed to have increased the propensity of large segments of the manpower pool to enlist. With ample supply, all Services--through operational cutting scores--achieved a large percentage of quality accessions (e.g., high school diploma graduates and/or AFQT scores at or above the 50th percentile). In response to such recruiting success, the Senate Appropriations Committee recently cut FY 1983 Defense personnel funds including recruiting incentives. No doubt the feeling was that with applicants banging on the Services' doors and quality accessions coming in, recruiting incentives would be unnecessary or at least a low priority item. Furthermore, Congress has set a 20 percent ceiling on below average personnel and has limited non-high school graduates to 25 percent in FY 1983. Such budget cuts and quality objectives are fine so long as other environmental factors such as high unemployment and low force requirements continue to positively affect accession quantity and quality. Rhetorically we may ask--what will happen to quality if numerical requirements increase sharply and/or the civilian labor force is not crippled by high unemployment? If it is true that history repeats itself, it is to history that we turn for the answer.

The selection process is a complex multivariate personnel management system. Although it is convenient to discuss environmental factors in isolation, they act as a unit. Despite this caveat, the present authors opt for convenience and primarily focus on mobilization status and youth unemployment rates in relation to changes in applicant qualification requirements and accession quality.

Response to War: Induction & Enlistment Standards During the Draft Years

Many changes in selection policies occurred during the draft period from 1940 to 1973. Mental standards for induction and enlistment varied mostly in response to the quantitative demands posed by World War II, the Korean Conflict, the Berlin Crisis, and the Vietnam War.

During most of the draft years, two sets of standards existed: one for inductees and generally a higher set for enlistees. Inductee standards were lower and tours of duty shorter for reasons of equity. All except the most untrainable must be eligible and accepted since the Selective Service could not justify picking only the cream of the crop to bear the brunt of compulsory service. Although the draft brought in many high aptitude personnel, it brought in marginal performers as well: therefore, shorter tours helped to prevent compromising the quality of future careerists and/or noncommissioned officers. Since it is to volunteers that the Services turn to first, even in times of war, draftees were used only to supplement the forces, particularly the Army with its large manpower demands and often inadequate market.

In times of war or national emergency, and to a lesser extent during peacetime recruiting shortages, the Army found it necessary to shift from qualitative considerations to quantitative demands. With each mobilization or manpower build-up, enlistment and induction standards were lowered to increase the size of the pool. Standards barring the induction of those with less than a fourth grade reading capacity at the initial phase of the World War II mobilization, for example, quickly proved too stringent. Concern over possible manpower shortages coupled with pressure from Southern Congressmen--whose constituents were being rejected at high rates--paved the way for a 10% illiterate quota system in August of 1942 (Wool, 1968). This was the Army's first experience of sacrificing quality for quantity.

The Navy's smaller manpower demands enabled it to avoid using inductees until 1943 when the Selective Service became the sole procurement agency and distributed illiterates to the Navy as well. From this time on, all Services were to be affected by the Army's quantity needs and quality problems, particularly in war.

Following the war (1946), reliance on the draft was reduced and higher peacetime enlistment standards prevailed. In order to forestall Army and Marine Corps manpower shortages under predominantly volunteer recruitment, the Selective Service Act of 1948 enabled the draft to become a peacetime procurement tool. This act established by law--for the first time--a specific minimum mental standard for induction which was higher than the World War II standard. Inductees were to be accepted if they achieved a standard score of 70 or better on the Army General Classification Test, corresponding to a percentile score of 13 on the AFQT. Even though the Army and Marine Corps needed the help of the draft, the standard was not set extraordinarily low, for they did not need "too much help" at this time.

In 1951, however, under the Universal Military Training and Service Act the minimum mental induction standard was lowered to the 10th percentile on the AFQT. This action was taken by Congress to broaden the manpower pool in light of the demands of the Korean Conflict. As in World War II the Army was the primary user of inductees and was saddled with a disproportionate amount of low aptitude personnel in comparison to the other Services.

To avoid a concentration of low quality personnel in the Army, DoD adopted a qualitative distribution policy from 1951 through 1958. This policy set all enlistment standards at the same level as inductees and required that each Service accept a specific percentage (quota) of personnel in mental categories I through IV. The quotas for low aptitude personnel (Mental Category IV) ranged from 27 percent to 12 percent of nonprior service accessions.

With strengths reduced following the Korean hostilities, the DoD imposed quotas were reduced and finally suspended in 1958. Not only were enlistment standards raised but in July of 1958 Congress authorized modifications to induction standards except in time of war or national emergency. This year marks the first time that supplemental aptitude tests were used along with AFQT criteria for screening inductees and enlistees, especially those scoring in the lowest acceptable aptitude category (i.e., Category IV - AFQT 10-30).

The period between 1958 and 1965 was a peacetime period somewhat disturbed by the Cold War and the Berlin Crisis (1962). Enlistment standards were set unilaterally by Service and generally ranged between AFQT 21 and 31 with varying supplemental test requirements. Between 1958 and 1963 induction standards required an AFQT 31 or AFQT 10-30 and standard scores of at least 90 in two or more Army Classification Battery aptitude composites. Those who failed were deferred from peacetime Service. In 1963 standards were raised further by adding a General Technical composite requirement of at least 80 for those in AFQT Category IV.

In November of 1965 Army and Marine Corps enlistee standards were set by DoD at approximately the same level as for inductees to assure a maximum input of volunteer enlistments (United States Congress, 1966). Previous supplementary aptitude test requirements were waived, for example, in the case of high school graduates with AFQT scores between 16 and 30. Two reasons can account for such a reduction in standards. First, volunteer enlistments may have been down in these two Services because of the sizable reduction in the national unemployment rate among males ages 18 to 24. In 1964 the rate was 9.7 while in 1965 it was only 8.1. Even more plausible, however, was the approaching U.S. involvement in Vietnam.

Table 1

Comparison of October 1965 and June 1966 Service Enlistment Aptitude Standards

	October 1965 (Pre-Vietnam)			June 1966 (During Vietnam)		
	AFQT	Aptitude tests	Education	AFQT	Aptitude tests	Education
Army-----------	31	-----------------	-------------	31	-----------------	-------------
	21-30	AQB-3----------	High school graduate	16-30	AQB-2----------	-------------
				16-30	-----------------	High School graduate
Navy-----------	31	-----------------	-------------	31	-----------------	
	21-30	-----------------	High school graduate	16-30	GT-80 plus 2 other AQB.	-------------
				21-30	-----------------	High School [illegible]ate
Quota--------	[a]	[a]	[a]	[b]	[b]	[b]
Marine Corps---	31	-----------------	-------------	31	-----------------	-------------
	21-30	AQB-3----------	-------------	16-30	AQB-2----------	-------------
				16-30	-----------------	High School graduate
Air Force------	AFQT-31---	1 area out of 4 in Air Force test, at percentile score of 40+	High school graduate preference.	[c]	[c]	[c]
	21-30	-----------------	High school graduate	-----------	-----------------	-------------

[a]12-percent group IV ceiling.
[b]5-percent group IV ceiling.
[c]No change.

Since Vietnam was never officially declared a war or even a national emergency, induction standards were not reduced to an AFQT of 10 which Congress called for under such conditions. Despite what Vietnam was called, numerical requirements increased and enlistment and induction standards were lowered. Table 1 compares the enlistment standards in effect just prior to the Vietnam build-up (October 1965) and those which operated in the midst of our involvement.

With the advent of the Vietnam war, test score and educational standards were lowered four times and DoD imposed quotas to accommodate the Army's numerical requirements and the fortuitous social program--Project 100,000. This program, as part of the President's War on Poverty, admitted low aptitude and previously rejected personnel into the military in order that they might learn useful skills. The goal was to admit 100,000 of these "New Standards Men" into the military annually. In addition to this general goal DoD established Category IV quotas ranging from 25 percent of Army accession to 15 percent of Air Force accessions. At least 50 percent of the Category IV quotas was to be met with "New Standards Men"; thus, men scoring in the AFQT range of 10 to 15 were brought into all Services. Towards the end of the Vietnam War, draft calls were reduced, Project 100,000 ended and standards were raised as plans for an all-volunteer force got underway.

Throughout the draft period the military's mobilization status and force strength requirements affected enlistment and induction standards. Although the specific standards varied, the pattern was essentially the same: with each manpower build up for war, standards were lowered and reliance on inductions increased to yield more accessions. Standards could be raised with the draft still operating to forestall Army and Marine Corps shortages. Although the Navy and Air Force had little trouble obtaining volunteers, (particularly with the draft stimulating enlistments), and could have maintained higher standards, DoD imposed quotas and lowered their standards so that the Army would not be saddled with all the low quality personnel. Generally, from 1940 to 1973, standards were affected by factors and policies internal to the military while external factors played more of a role once the draft ended.

Quality Objectives in an All-Volunteer Environment

The years 1972 and 1973 are known as the transition period to the All-Volunteer Force (AVF). With declining draft pressure and abolition of DoD quotas, the Services began to shift their entry standards and ceilings on Category IV and non-high school graduate personnel in order to find the best quality mix that their individual markets would support (Lee & Parker, 1977). In their efforts to maximize quality during this time when the market was changing, the Army, Navy, and Marine Corps sometimes experienced recruiting shortfalls. Quality objectives were then lowered and standards adjusted in response to shortfalls. The Air Force set relatively high standards and were able to maintain them and even flourish under the free market of the AVF due to its more favorable image, adequate supply, and lower numerical requirements than the Army and Navy.

With the draft gone as a peacetime procurement tool, the Services could no longer afford to set standards and objectives unrealistically high since inductees and draft motivated enlistees were no longer available to fall back on. Through trial and error standards were set in light of manpower availability as well as quality demands. Early in 1973 the Marine Corps, for example, required a General Technical (GT) composite score of at least 80 and standard scores of 90 on two additional aptitude composites for all applicants with AFQT scores between 21 and 49. In order to increase supply, all composite requirements were dropped for high school graduates. In addition, GT requirements were later dropped for non-graduates scoring between AFQT 31 and 49 and graduates scoring between the 21st and 30th AFQT percentiles.

Table 2

1982 Male Non-Prior Service Enlistment Aptitude Standards
(Required Operational Score on ASVAB 8 - 10) by
Educational Level

Service/Education	MINIMUM STANDARDS AFQT Percentile Score	Aptitude Composite[a] Standard/Percentile Score
ARMY		
H.S. Diploma Graduate	16	85 on Any 1
GED	31	85 on Any 1
Non-H.S. Graduate	31	85 on Any 2
NAVY		
H.S. Diploma Graduate	17	b
GED or CPT	31	b
Non-H.S. Graduate	28	b
MARINE CORPS		
H.S. Diploma Graduate	21	GT[c]=80
Non-H.S. Graduate (Including GED)	31	GT[c]=100
AIR FORCE		
H.S. Diploma Graduate	21	G[d]=30; MAGE[e]=120
GED	50	G[d]=30; MAGE[e]=120
Non-H.S. Graduate	65	G[d]=30; MAGE[e]=120

[a]Minimum composite scores are expressed in terms of standard scores for the Army, Navy, and Marine Corps. Percentile scores are used in the Air Force.

[b]No minimum requirements for enlistment.

[c]General-Technical ASVAB Composite.

[d]General ASVAB Composite.

[e]Mechanical, Administrative, General, & Electronics ASVAB Composites.

Toward the late 1970s minimum AVF enlistment standards were set at levels which were practical for each service. With minor adjustments along the way, minimum standards evolved into those of today. As shown in Table 2, each service has a unique set of minimum AFQT and aptitude composite standards. For all Services, these requirements are more stringent for non-high school graduates and GED recipients (because of their higher first term attrition rate) than they are for high school diploma graduates.

While minimum standards do not preclude the enlistment of Category IV or non-high school graduate personnel, they are the least preferred group of accessions. Good quality, on the other hand, is generally defined as high school diploma graduates scoring in AFQT categories I-IIIA (i.e. AFQT 50 through 99). When market conditions are favorable, the Services often set operational cutting scores and quality objectives at levels higher than the minimum standards, thus pursuing the more desirable candidates. Environmental factors external to the military have played an increasing role in the military selection process since the inception of the AVF. There is a strong indication, for example, of an inverse relationship between the nations overall economic health and the ability to attract an adequate number of well-qualified youth into Service (Toomepuu, 1981; Philpott, 1982). When youth unemployment is low and competition with the private sector is fierce, the Service recruiters tend to enlist individuals as they apply, thus bringing in more individuals who score closer to the minimum standards. When unemployment is high the Services are afforded the luxury of choice and can enlist more preferred quality personnel. Although it is difficult to state what the actual cutting scores are for each branch, it is possible to see their effect on the quality of accessions.

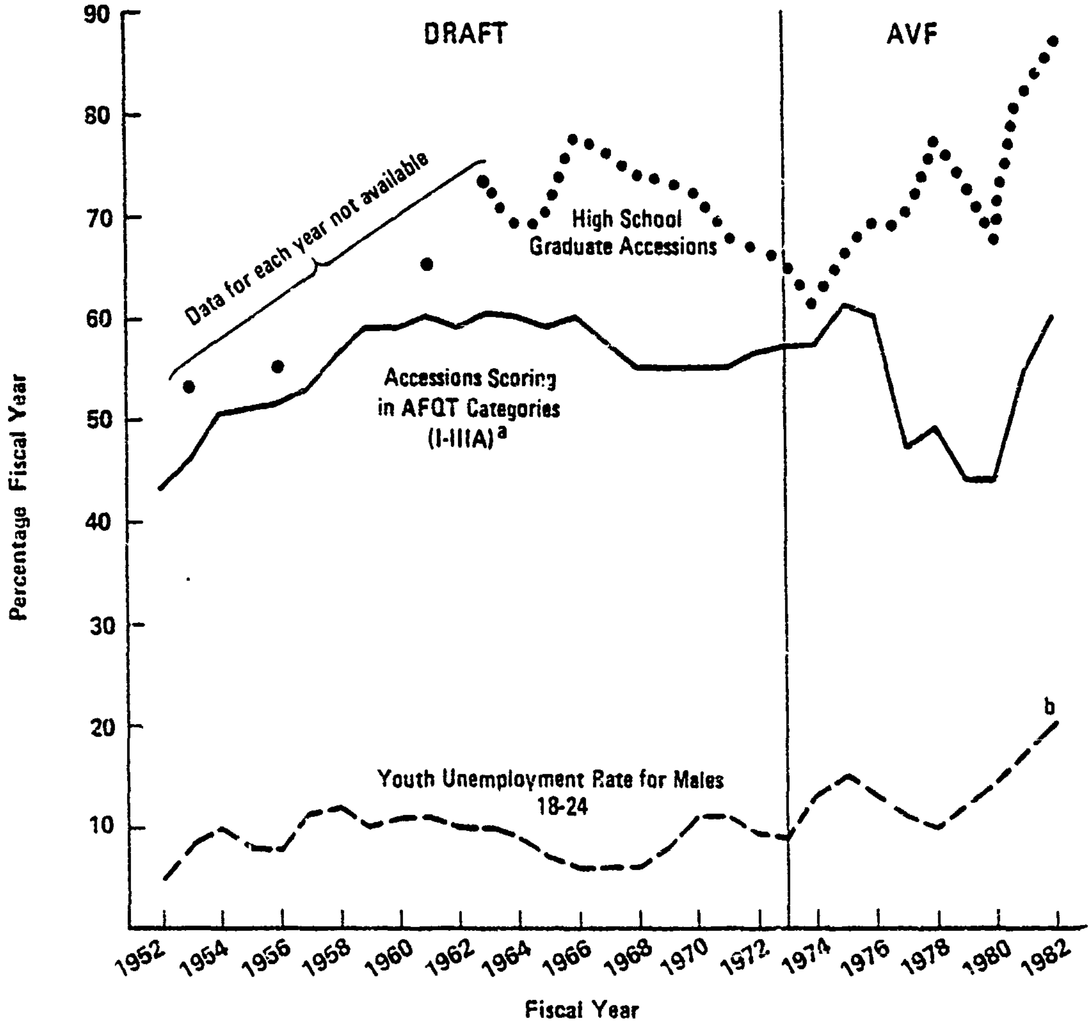


Figure 2. Quality of Male Non-Prior Service Accessions, as Measured by AFQT Categories I-IIIA and High School Graduation, in relation to the Unemployment Rate For Male Youth Ages 18-24. (Total DoD, Fiscal Years 1952 through 1982)

a. Categories I-III A correspond to scores at or above the 50th percentile on AFQT.
b. The Youth Unemployment Rate was calculated from data provided by the Bureau of Labor Statistics, Current Population Survey.

Figure 2 tracks the quality of accessions and youth unemployment rates from fiscal year 1952 through 1982. While there is no discernable pattern between quality and civilian unemployment during the draft period, a clear relationship does exist under the AVF. Quality shifted prior to 1973 mostly in response to force strength requirements. Between 1966 and 1971, for example, requirements for Vietnam and Project 100,000 led to a decrease in the percentage of above average AFQT and high school graduate accessions. Economic conditions appear to be irrelevant until the AVF was firmly established. Since then the Services have been playing the manpower market--maximizing their intake of Category I-IIIAs and high school graduates when unemployment rates rise. From these fluctuations in quality one can assume that the Services have been flexible in their application of minimum enlistment aptitude standards, adjusting them upward when conditions permit. In trading off quality for quantity, it appears from the AVF side of Figure 2 that aptitude level is sacrificed before education. Under unfavorable market conditions the Services continue to pursue high school graduates, but increase supply by enlisting them with scores close to, or at, the minimum standards.

Report Implications

While recognizing the complexity of the military personnel procurement process, this report has indicated that environmental events must be considered in setting aptitude selection requirements. Events both internal and external to the military act as warning signs which may lead to a change in selection standards and daily recruit quality objectives. If we assume that the All-Volunteer Force will continue to operate in the future then external factors, such as the unemployment rate, will continue to have a strong impact upon the quality of accessions.

The time is ripe for evaluating enlistment standards and quality objectives. Chances are, unemployment rates will descend in the near future. Recent history has shown us that with active competition from the civilian labor market, the Services (particularly the Army) will tend to experience recruiting shortages and react by lowering operational cutting scores. The current high cutting scores and accession quality may be affected by other factors as well.

In addition to the negative impact of expected lower unemployment rates, recruit supply may suffer from a decline in the size of the prime manpower pool (i.e. male youth ages 18-23). Although technological advances will continue along with preferences for recruits who are high school graduates in AFQT Categories I to IIIA, the Services may be forced to select at their minimum standards. Depending upon the Severity of the supply-demand ratio, it is possible that minimum standards might also be affected. DoD is making some preparations through its investigations of less preferred segments of the manpower pool such as non-high school graduates.

Finally there is one more implication offered. From the many standards and cutting score changes, it appears that the "quality" sought is a function of the "quality" available. Minimum standards are based, to a large extent, on Service preferences, market conditions, and training ease. With the present efforts by the Services and DoD to link aptitude standards to actual job performance we may indeed be headed towards a change in standards. Hopefully research efforts may reduce "demands" and pave the way to more efficient utilization of the personnel that are able to be recruited.

References

Lee, G. C., & Parker, G. Y. Ending the draft - The story of the all volunteer force (FR-PO-77-1). Alexandria, VA: The Human Resources Research Organization, April 1977.

Philpott, T. "Joblessness linked to recruiting highs" Air Force Times, October 11, 1982, p. 7.

Toomepuu, J. Soldier capability - Army combat effectiveness (SCACE). Volume I Main Report (ACN64024). Fort Benjamin Harrison, IN: U.S. Army Soldier Support Center, USAREC, April 1981.

United States Congress, Review of the administration and operation of the Selective Service System (No. 75). Hearings before the Committee on Armed Services, House of Representatives, Eighty-Ninth Congress, June 1966.

Wool, H. The military specialist: Skilled manpower for the Armed Forces. Baltimore, MD: The Johns Hopkins Press, 1968.

AD P000889

Differences on Personality Measures Related to Recruit Attrition

Lieutenant-Colonel D.A.L. Lefroy
Canadian Forces Personnel Applied Research Unit (CFPARU)

Introduction

Cotton (1974) in his demographic studies has indicated that major problems relating to recruiting and attrition will face the Canadian Forces over the next decade. Not only will the recruiting base be smaller but the potential recruit will be better educated and with a different set of expectations. The present study attempts to shed further light on the phenomena of attrition by focussing on attrition during recruit training. This choice was influenced by the studies of Porter and Steers (1979) who indicate that the initial period of membership in an organization is the most critical as that is when most attrition occurs, and by VanMaanen and Schein (1977) who point out that a large number of studies demonstrate that early organizational experiences impact on one's later organizational behaviour.

However, recent reviews of the research literature on personnel turnover indicate that measures of personality, interests, and intelligence do not reveal what could be considered a consistent relationship with turnover across situations (Muchinsky and Tuttle, 1979; Mobley, Griffeth, Hand, and Meglino, 1979). Consistent cross-situational predictors were found to be personal predictors (age), attitudinal predictors (job satisfaction), and work related predictors (leadership). Mullin (1980) in reviewing these summaries and other research concluded with respect to the studies in attrition:

1. The knowledge of such "explanatory fiction" as "Job Satisfaction" or "Organizational Commitment" may be of descriptive or predictive value but adds nothing to the knowledge of the dynamics of attrition.

2. The level of aggregation has not been sufficiently dealt with in analysing attrition. That is to say, a macro organizational perspective tends to mask important differences at the sub unit level.

3. Grouping of all types of leavers into a single category within the stay/leave criterion may mask the potential predictive value of sub classes within the criterion group.

In the present attempt to determine whether personality variables relate to recruit attrition in the Canadian Forces, and what might underlie attrition, these observations were taken into account. Consequently, the main focus was placed on the within squad interaction between the NCO, the recruit and the recruit peer group. Also, several discrete and composite attrition categories were utilized.

To obtain useful answers to the questions raised relating to personality variables and the dynamics of their effects on vocational change, one must have a theory of vocational development that is broad enough to incorporate work values, interests, or beliefs, that is researchable, and is pragmatic enough to be useful to military career counsellors. The work of Holland (1973), to a large extent, satisfies these requirements by presenting both a logical and an empirical framework. In his personality-environment congruence hypothesis, Holland (1966) considers vocational achievement, satisfaction and stability to be related to the congruency between one's personality and the vocational environment largely composed of other people. Therefore, it was hypothesized in this study that recruits whose personality measures were similar to those of the squad NCO and to the largest personality grouping in the squad (modal) would show lower attrition than recruits whose personality measures were different. It was also hypothesized that work values were related to attrition. A third hypothesis was that in a high constraint, high discipline environment, with structured leadership, recruits with an External Locus of Control would have lower attrition.

Method

In order to reliably measure different facets of personality that would likely relate to attrition, a number of personality measures were administered to all recruits on arrival at the Canadian Forces Recruit School, Cornwallis, and to their squad NCOs. These included Holland's Preference Inventory (VPI), Levenson's Locus of Control Scale (LCS) and Super's Work Values Inventory (WVI). Holland's VPI is a personality linked measure of vocational interest (Holland 1966) which implies that personality has a determining effect on choice of vocation. The Locus of Control Scale (Rotter 1954) is the measure of generalized expectancy or belief in the connection between one's behaviour and the occurrence of outcomes, thus affecting one's adaptation to life events. For example, Internals believe that their behaviour is responsible for reward and punishment, while Externals (C) attribute reward and punishment to fate or Chance and Externals (PO) attribute both to the action of Powerful Others. With respect to his WVI, Super (1957) implies that one's value system is a significant variable in the selection of a career. Thus, life values find expression in work. In all, 15 Work Values are measured.

During the periods 15 October 1979 to 18 November 1979 and 27 January 1980 to 3 March 1980, a total sample of 1306 English speaking male recruits ranging in age from 17-23 undergoing an eleven week basic training course at Canadian Forces Recruit School, Cornwallis were available of which 1070 were tested (on arrival) and 980 were used; (90 were released for reasons relating to purely medical, social or learning problems). The 30 squad NCOs responsible for this sample of recruits (comprising 41 squads ranging in size from 21 to 44 recruits) were also part of the sample. Fourteen of the NCOs commanded two different squads and thirteen commanded one squad only. Three acted in an assisting capacity only. All were experienced instructors from the combat arms.

Eight single and three composite attrition categories were utilized:

Category	Definition
0	completed training - not recoursed
2	failing course - requested release - granted
3	passing course - requested release - granted
4	failing course - learning ability
5	released - medical (physical only)
6	released - social (theft, homosexuality, chronic drug use)
7	poor performance - recoursed and failed again
9	poor performance - recoursed and passed
F-I	failed to adjust to initial squad (2, 3, 7, 9)
F-II	eventually left forces because of adjustment (2,3,7)
F-III	designated by squad NCO as failing (2,7,9).
P-I	Pass (0)
P-II	Pass (0+9)
P-III	Pass (0+3+9)

Table 1 VPI Types Related to Attrition From Forces

Personality Types (grouped)

I	II
R (realistic)	E (Enterprising)
C (conventional)	S (Social)
I (Investigative)	A (Artistic)

	I	II	
Pass II	659	192	851
Fail II	70	36	106
	729	228	957

$\chi^2 = 6.369$ df = 1 Sig = .0132 $p < 05$

Table 2 LCS Recruit-Environment Congruency within Squads Related to Attrition (recruit LCS type similar both to NCO + squad mode)

	No	Yes	
Pass I	661	134	795
Fail I	166	19	185
	827	153	980

$\chi^2 = 4.45240$ df = 1 Sig = .0349 $p < .05$

Table 3 — Differences in Mean Scores on Levenson's LCS Sub Scales I, E_c, E_{po} between Recruits who adapted to their Squad and those who did not

Sub Scale	Cat	M	SD	N	df	F	Sig
Internal	Pass I	35.977	6.639	795	1	1.315	.2518
I	Fail I	35.357	6.595	185			
External Chance	Pass I	17.355	7.557	795	1	4.035	.0449*
E(C)	Fail I	18.600	7.758	185			
External	Pass I	22.547	8.686	795	1	0.017	.8977
Powerful Others E(PO)	Fail I	22.638	8.423	185			

* $p < .05$

Table 4 — Significant Correlations Between a Subset of Super's Fifteen Work Values and Six Categories of Attrition

V	2	3	7	9	F-III	F-I
CR	-.030	-.005	-.062	-.047	-.068*	-.063*
Ma	-.081*	.009	-.012	-.037	-.069*	-.059
Su	.067*	.036	-.009	.017	.042	.049
WL	-.034	.066*	-.052	-.070*	-.081*	-.053
ER	-.021	.038	-.055	-.094**	-.090**	-.070*
IS	-.019	.010	-.073*	-.041	-.063*	-.053

$p < .05$* — F-III = (NCO designated failures)
$p < .01$** — F-I = (adaptation to squad)

CR = Creativity
Ma = Management
Su = Surroundings
WL = Way of Life
ER = Economic Returns
IS = Intellectual Stimulation

Results and Discussion

In Table 1, due to the small number of recruits with C, E, S and A personality characteristics the data were collapsed using Holland's Hexagonal model which groups personality types according to their similarity (Holland 1973). As a group the Enterprising, Social and Artistic types showed significantly higher attrition from recruit training than the Realistic, Investigative, Conventional types, who by their numbers constituted an I, R, C environment, thus supporting Holland's personality-environment congruence hypothesis as a theory of career change.

In Table 2, Holland's personality-environment hypothesis is again supported. However, only when there was an internally consistent environment was attrition significantly less, that is, when recruits shared perceptions on the squad reinforcement contingencies with both the squad NCO and modal group type within the squad. It is noteworthy that ten of the sixteen congruent squads were Internals. Cook et al (1980) found when accounting for differences in attrition between platoons during Marine Corps Training the Locus of Control was found to be significantly related to attrition. It was found that a change in the Internal direction occurred in the low and medium attrition platoons while a change in the External direction occurred in the high attrition platoons. The authors propose that different training environments have a mediating effect. A consistent Internal environment would maximize this effect.

In Table 3, those who do not adapt to their initial squad tend to score significantly higher on the External Chance sub scale with no differences on the I or E_{po} sub scales. The studies from which the hypothesis was formulated that Externals would adapt better to the high constraint, high structure, high discipline military environment (Parent et al 1975; and Wolk 1976) did not control for local reinforcement contingencies. As indicated by the Cook et al (1980) study reinforcement contingencies independant from these three factors could lead to shifts in perception of control with consequent effects in adaptation. Those scoring high on Ec would require a greater shift and consequently would be less likely to adapt.

In Table 4, the largely negative correlation between the work values; Creativity, Management, Economic Returns, Intellectual Stimulation and Attrition would be what one might expect in a largely Realistic group of individuals with their "blue collar" orientation. Two values correlate positively with attrition, Way of Life and Surroundings. Those who perform well but request release tend to value Way of Life highly. Those who perform poorly (failing) and request release, value Surroundings highly. It is interesting that in spite of disparate data sets, military people in the United Kingdom, United States and Canada commonly express unhappiness with pay and lifestyle (Wiskoff and Mutlock 1980). However, in view of the negative correlation with attrition it would seem that pay is not an important consideration at the Recruit level.

The results of this study would suggest that there is a significant relationship between personality as measured by VPI, WVE and LCS and attrition from the Canadian Forces Recruit school. The evidence suggests an interaction effect between the individuals personality and the environment as defined by other personality types. Also, the approach confirms the usefulness of a micro organizational design and the use of multiple dependant variables.

References

Cook, T.M., Novaco, R.W., & Sarason, I.G. Generalized expectancies, life experiences, and adaptation to Marine Corps recruit training. Report AR-002. Office of Naval Research, Arlington Va. April 1980.

Cotton, C.A.U. Unskilled male recruits for the Canadian Forces: Trends and projections to 1985. Canadian Forces Personnel Applied Research Unit Report 74-5, Oct 74.

Holland, J.L. Psychology of Vocational Choice. Waltham: Blaisdell, 1966.

Holland, J.L. Making Vocational Choices: A theory of careers. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.

Mobley, W.H., Griffeth, R.W., Hand, H.H., & Meglino, B.M. Review and conceptual analysis of the employee turnover process. Psychological Bulletin, 1979, 86, 493-522.

Muchinsky, P.M., & Tuttle, M.L. Employee turnover: An empirical and Methodological assessment. Journal of Vocational Behavior, 1979, 14, 43-47.

Mullin, E.E.D. Utility of futility in attrition research. The P.S.O. Forum, 1980, 6, 92-109.

Parent, J., Forward, J., Centre, R., & Mohling, J. Interactive effects of teaching strategy and personal locus of control on student performance and satisfaction. Journal of Educational Psychology, 1975, 67, 764-769.

Porter, L.W. & Steers, R.M. Organizational commitment and personnel attrition. San Diego, California, Office of Naval Research, 1979.

Rotter, J.B. Social Learning and Clinical Psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1954.

Super, D.E. The Psychology of Careers. New York, Harper and Row, 1957.

VanMaanen, J., & Schein, E.H. Toward a theory of organizational socialization. Annual Review of Research in Organizational Behavior, Vol I, 1977.

Wiskoff, M.F., & Matlock, M.M. Reasons for separating: A comparative analysis. Technical Panel in Military Manpower Trends (UTP-3) San Diego, California. Navy Personnel Research and Development Centre. June 1980.

Wolk, S. Situational constraints as a moderator of the locus of control-adjustment relationship Journal of Consulting and Clinical Psychology, 1976, 44, 420-427.

HUMAN PERFORMANCE REQUIREMENTS IN C^3I SYSTEMS AND THEIR IMPLICATIONS IN SYSTEM DESIGN

Dr. Leslie Lewis TRW/DSG
Melinda Copeland TRW/DSG

Human performance in Command, Control, Communications and Intelligence (C^3I) systems is an area of increasing concern among system designers. This interest in the human dimension arises from the recognition by system designers that the operators are the critical links in the system; they perform the analytical work of defining the battlefield and, based on that analysis, provide information on which commanders' decisions are made.

To do this work the C^3I operator uses a variety of cognitive skills or mental processes to analyze data presented to him by the system. The system provides analytical tools to assist him in this work. It is critical to system design that the cognitive performance requirements be clearly defined in order to identify optimal data presentation methods and formats, as well as software tools and aids for the operator.

Unfortunately, the identification of the cognitive processes essential to performing intelligence analysis, and the translation of these processes into system requirements, is difficult because cognitive processes are not readily definable into quantitative and testable units. TRW has begun to address these issues in several program-related Independent Research and Development (IR&D) activities. This paper concentrates on some of the work done in the user interface area and the design implications specific to C^3I systems.

Cognitive skills are information processing techniques used to restructure one's knowledge of a situation. The term cognitive skills is often used synonymously with problem-solving skills. Human information processes vary widely in their complexity. In most tactical intelligence environments, combinations of skills are used by the most successful analysts to perform threat analyses, for example. Such skills as recall and recognition memory are classified as low-level cognitive skills, while deductive and inductive reasoning are high-level, sophisticated skills. TRW research for several C^3I systems (Tactical Computer System/Tactical Computer Terminal (TCS/TCT), Corps Support Weapon System (CSWS) and Battlefield Exploitation and Target Acquisition (BETA)Testbed), has revealed that there are thirteen types of cognitive tasks required for operators to perform various battlefield analyses. These tasks and associated subtasks are presented in Table I. All of these tasks are performed in some form and in varying degrees of complexity for all the systems analyzed. How the operator prioritizes, performs and completes the tasks is determined by the message and the medium in which the information is received, processed and transmitted by the operator. In the systems studied, five message-media types, i.e. ways in which data were presented to the users, were found: operational paperwork, manuals and technical documentation, hardcopy messages,

TABLE I. REQUIRED COGNITIVE TASKS OF A C^3I OPERATOR/ANALYST

	COGNITIVE TASK	SUBTASKS COGNITIVE AND REASONING SKILLS
1.	Determine Requirements/Criteria	o Comprehend concepts o Formulate new requirements o Translate abstract ideas into meaningful requirement criteria o Integrate requirements into priorities
2.	Plan Action Sequence	o Determine and sequence mental and physical actions o Plan simultaneous computer and mental processes o Change mental strategy flexibility
3.	Assess Situation	o Comprehend global information o Match strategies to the appropriate problems o Form appropriate concepts
4.	Store and Retrieve Information from Computer	o Use all equipment efficiently o Know data base o Know data sources
5.	Translate Symbols into Information	o Recognize pattern o Transform pattern information into usable data o Translate abstractions into trends and patterns
6.	Reason Inductively	o Macro-to-micro reasoning
7.	Reason Deductively	o Micro-to-macro reasoning
8.	Generate Hypotheses	o Synthesize data o Recognize inconsistencies o Fill in gaps or aborts
9.	Formulate Probabilities	o Develop alternate hypotheses o Develop probabilities of hypotheses
10.	Test Hypothesis	o Relate changes in tactical situation to hypothesis o Change hypothesis o Formulate new hypothesis
11.	Visualize Dimensions of Time	o Ability of analyst to see entire situation in snapshot
12.	Synthesize Data into Comprehensive Whole	o Present total data o Eliminate perspective misconceptions o Transform tactical data into usable information
13.	Debriefs in Order to Replan and Reassess	o Conduct lessons learned evaluation o Amend 2 and 3 above into an updated protocol

verbal interactions and computer displays. The type and complexity of the message-media contribute significantly to the difficulty of the operator's job and frequently require the most sophisticated cognitive skills. Preliminary research on the above systems indicates that critical performance deficiencies exist in at least seven of the thirteen cognitive skill categories, with severe deficiencies in four: Formulate Probabilities, Test Hypotheses, Visualize Dimensions of Time and Synthesize Data into Coherent Wholes. Lack of these cognitive skills is either ignored or treated inadequately in the system design process. For example, it is vital that operators be able to visualize dimensions of time to convey an accurate picture of the battlefield to a commander. Yet, we found that little or no data were provided to assist the operator in this task.

Two problems exist in a poorly defined C^3I user interface. First, most operators are critically deficient in one or more cognitive skills. Secondly, the computer graphics and alphanumerics present the data in a format which hinders the development of the skill. The skill remains undeveloped, and the operators rely on supervisors or more experienced personnel for assistance, resulting in an inefficient distribution of the workload to the better trained operators.

TRW is convinced that a rigorous systems engineering activity is the key element in defining an optimal user interface. The term user interface refers to the user's operational environment. This environment is composed of many interfaces, all of which affect user performance. Since these interfaces cut across all major subsystems, the definition of the user interface must accompany other system- and subsystem-level definition efforts. There is a myriad of technologies that affect the user interface. Among the most critical are: (1) requirements analysis, (2) operational thread analysis, (3) user command and query languages, (4) data base management systems and (5) simulators and testbeds.

The requirements analysis within most design efforts concentrates on performance requirements and bounding the system from a hardware and software perspective. Human performance and user interface issues are either not dealt with or are described within the context of human factors. This orientation establishes requirements concerning environmental factors, keyboard layouts and the location of knobs. Though these issues are important, they do not define key system level elements nor do they drive system design. We are finding that when front-end analyses consider the user interface, rather than strictly hardware oriented human factors issues, interfaces among system elements which are critical to the user can be determined early. The user interface can then be incorporated into the overall system design, and in many cases becomes a subsystem. As the system requirements evolve into B level development specifications and C level product specifications, detailed relationships between the user interface, hardware and software sub-systems become clear.

Operational thread analysis is a part of the front-end systems engineering that is critical to and goes hand-in-hand with requirements definition and functional analysis. We have performed operational

thread analyses on several systems within TRW. It is a time consuming, iterative process but has been extremely beneficial from the perspective of user interface design. It supports functional allocation and allows systems engineers to define functions which should be automated. It also simplifies the identification of user interface technologies critical to the system, e.g., decision aids, command language structure and software tools. The thread analysis begins with the identification of processes currently being performed in the environment studied. These consist of processes currently performed manually as well as processes which are already partially or fully automated. Part of the operational thread analysis, then, is a definition of what functions the user performs, and most importantly, how the system supports the user. Apart from the user interface design benefits, operational thread analyses allow systems engineers to see how the system will operate across functions and subsystems, i.e. it demonstrates by example the system level operational concept.

The two user interface technologies which have the greatest effect on human performance, and which depend heavily on front end analyses in order to be treated appropriately in the system design, are user command or query languages and data base management systems.

Command languages are among the poor relations of the computer language family. At the present time, there are signs of rapid growth in both the importance and power of command languages. Their growth stems from the proliferation of terminals providing access to single systems and to networks. The people who are using command languages are not always specialists/programmers and do not want to learn a command language per se; rather, these people want to feel at ease in using and interacting with a terminal. The response to the expansion of computer technology has been that command languages began to resemble programming languages. Frequently the very need for a command language was defeated by the fact that it was written by programmers who were sensitive to programming needs but often lost sight of why they were writing the language -- to serve the user better.

The current trends and structures of command languages are heavily influenced by networks. Here, the critical user interface design principle is uniformity of the interface across the network. We are finding that part of this uniformity in C^3I systems is dependent upon a command language and structure that is dialogue-oriented and maps to a menu structure. Our experience with the Army and Air Force shows that defining a language which is consistent with the user's mental model of task operations, and has easily comprehensible semantic and syntactic structures, is necessary to support users in invoking and utilizing system functions. We feel that such a language can be developed for military C^3I systems, and that it will provide adequate fault tolerance and fast response times (both of the system and the user). The success of the language hinges, however, on thorough and accurate operational analyses.

Another issue related to the user command language concerns the system manager command language. The structure of this language is a problem since it must be both a programming language as well as a medium for interactive dialogue. It is not clear whether or not this language should be

mutually exclusive from the user language. Resolution of this question is dependent upon the definition of the system manager's role and tasks for a particular system.

The semantics and syntax of a command language are inextricably related to the data base structure. The capabilities of command languages to handle the relationships between data and computer programs which use that data are currently based on data bases structured via conventional file methods. This is inadequate and inappropriate for DBMS data bases. A DBMS data base is desirable because it supports a more flexible and transparent user interface. For example, conventional file structure data bases do not allow data element naming at the file level; or put another way, a file, record and field cannot all be named in the same command, but require three separate transactions. However, data element naming in a DBMS system may be done in one transaction as with <filename>. <recordname>. <fieldname>. A disadvantage with this scheme is that naive and casual users are intimidated by having to know complex storage attributes such as file/record/field structures. A system is needed with multiple interfaces for a variety of users. A top-down, layered, abstract machine approach to computer system design will achieve this. In this concept the outermost layer presents the simplest interface for the most inexperienced user. The simplicity may be achieved by presenting only limited capabilities, or by presenting the same capabilities available to the experienced users, but in a way tailored to the naive user. Each layer of the machine should have all the capabilities of the previous layers plus capabilities specific to that layer. What these capabilities are depends on the application environment of the system and who the end users will be.

The relationships between command languages and data bases will become more important and complex as the diversity of C^3I systems increases. The driver in this relationship must be the consideration of the mental model of the user both in terms of how he perceives system operations and what cognitive processes he uses to perform his job.

Finally, a critical element in the validation of a user interface is the building of a simulator. Accurate human modelling is difficult for dynamic, real-time C^3I systems, so simulation with the user in the loop is necessary to closely approximate system operation.

There are four major types of simulators. Test driver simulators provide simulated messages and inputs for test purposes. Performance analysis simulators are used to determine the quantitative parameters of the system, usually in terms of throughput. These simulators test whether or not the system requirements have been met. Training simulators are used to train operators to use a system where on-the-job training would be cost or time prohibitive.

User interface simulators, often called testbeds, are the most applicable type of simulator for this discussion. These testbeds generally perform qualitative assessments of the system, e.g., determining whether or not incoming data are sufficient for human decision-making tasks. These testbeds are also used to validate operator and/or system level operational concepts. This is done by validating the operational thread analyses. Once valid threads are defined, user interface elements and technologies, which were missed in the top level analyses, often become obvious. The

simulation and validation of operational threads is an ongoing, iterative process performed in concert with increasingly detailed system design.

We are seeing that the term man-machine interface (MMI) is deceptive. A critical step in system design is the definition and bounding of interfaces to systems and subsystems. The MMI is bounded too narrowly; it is too often limited to the concept of "man-at-the-console". As shown earlier in this discussion, the user interface implications and requirements go far beyond that concept.

Published References

Beech, D., ed. Command Language Directions, Computer Science Press, Inc., 1980.

Blanc, R. P., and Cotton, I.W., eds. Computer Networking, IEEE Press, 1976.

Human Performance Deficiencies and Requirements in C I Systems - Appendix I to 1980 Man-Machine Interface IR&D, 31 December 1980.

Lano, R. J., A Technique for Software and Systems Design: TRW Series on Software Technology, Volume 3, North-Holland Publishing Company, 1979.

Lewis, Leslie K., Military Avionics Maintenance Performance Requirements, 1985-2000, 1979.

Ullman, J. D., Principles of Database Systems, Computer Science Press, Inc., 1980.

ORGANIZATIONAL AND NON-ORGANIZATIONAL INFLUENCES ON ATTITUDES TOWARDS WORK

Robin I. Lissak, University of Illinois
Captain H. Mendes, Canadian Forces Personnel Applied Research Unit (CFPARU)
E. Allan Lind, University of Illinois

Research and theory on attitudes towards work has long been based on the view that the rewards obtained by working are a major determinant of attitudes. "Exchange" theorists (e.g., Adams, 1965; Homans, 1961) have gone so far as to point to relative or absolute level of reward as the central cause of worker satisfaction or dissatisfaction with work. We will describe here a study of attitudes in the Canadian Forces that suggests that this emphasis on the level of outcome is misplaced. In particular, we will present data that shows that an individual's satisfaction with the organizational procedures used to allocate outcomes is at least as important as his satisfaction with the outcomes themselves in affecting overall job satisfaction.

Many studies have examined the relationship between the magnitude of reward associated with a job and the individual's attitudes towards work. A consistent finding in this research has been that there exists a positive, but weak, relation between reward and attitude. That is, higher rewards are associated with more positive attitudes, but most of the variation in attitudes cannot be explained by reward magnitude. The weakness of the reward-attitude relation has prompted theorists to seek more sophisticated notions of reward, which seek to allow for individual differences in the meaning or value of rewards. But even when these more sophisticated models of reward are used to predict attitudes, there is much variation in work attitudes that is not explained.

We undertook the present study in order to test whether at least part of the unexplained variation in work attitudes could be accounted for by the organizational procedures that link rewards to behavior. Prior to the present study there had been almost no work on the general effects of organizational procedures on job attitudes. We were nevertheless persuaded by a growing body of literature in social psychology that such effects do exist and that procedures might well explain the "missing" variation in job attitudes. The social psychological studies had shown that the procedures used to resolve disputes, assign grades, or make political decisions had substantial effects on the attitudes of those affected by the procedures. These effects occur independently of the outcome of the procedures; even if an individual receives a poor outcome, the use of a particular procedure that is seen as fair makes the outcome more palatable and produces more favorable reactions to the organization. Similarly, the use of a procedure that is seen as fair makes favorable outcomes even more satisfying than they would be otherwise and, again, produces more favorable reactions to the organization. We reasoned that such procedural effects might well occur in the Canadian Forces and that they might play a major role in affecting, for better or worse, the attitudes of individuals towards their jobs and towards the Forces.

In deciding to examine the role of procedures in affecting job attitudes, we were also influenced by a practical consideration. Although it would be of some academic interest to discover any factor that influences job attitudes, it would be of considerable practical importance if that factor was something under the control of the Forces. Procedures, because they are promulgated by the Forces for the Forces, are more easily modified than are other potential influences on attitudes. The line of reasoning we followed went as follows: suppose we find that job attitudes are affected both by an individual's satisfaction with the rewards associated with military service and by his satisfaction with the procedures used to allocate those rewards. Suppose further that we wish to improve attitudes to achieve our overall goals in the Forces. There is relatively little we can do within the Forces to accomplish major changes in the level of rewards our people receive, but there is much more freedom of action in changing the procedures used to decide how those rewards are to be distributed. If it is the case that procedures play a major role in work attitudes, we need to know that this is so, in order to know what options exist for change and improvement.

The Canadian Forces, like any military organization, uses procedures to govern and regulate its actions. We are concerned here, in particular, with the procedures that govern the relationship between the Forces and its members. Like most Armed Forces, we have procedures that govern our performance evaluation process, procedures that govern our promotion and pay decisions, procedures that control posting decisions, procedures that control retirement benefits, and procedures that allocate the work resources our people use to do their jobs. These procedures are appropriately judged on the basis of many criteria besides whether they engender positive attitudes, but their potential effect on attitudes cannot be ignored. Most of us have seen procedures that might have been a good idea with respect to the quality of decisions they produce but that engendered such resentment that both general morale and the functioning of the procedure itself suffered. A major purpose of our study was to determine how extensively attitudinal reactions to procedures affect overall attitudinal reactions to one's work.

The social psychological research and the theory it has generated suggests that procedures engender positive attitudes to the extent that they are seen as allowing those affected by the decision some "say", some input of information prior to the decision. The previous research also shows that the favorability of an individual's reaction to a procedure could be measured by soliciting ratings of his "satisfaction with procedure" or ratings of the "fairness of the procedure". Another measure that has been shown to be closely tied to evaluations of procedures is an individual's ratings of how satisfied he is with his "treatment' by an organization. The study described below collected ratings on each of these groups of measures of reactions to procedures, and it collected comparable ratings on measures of reactions to several classes of rewards or benefits that one receives as a member of the Canadian Forces. In addition, data was collected on general attitudinal reactions to work in the Forces. Our major hypothesis was that measures of reactions to procedures would be at least as important as the measures of reactions to outcomes in predicting job satisfaction.

Method

The data reported in this study was collected as part of a pilot study on attrition and retention in the Canadian Forces conducted by the Canadian Forces Personnel Applied Research Unit. The data was collected at four military bases in Canada from 262 military personnel. A complete description of the survey, sample and method is provided in Lissak and Mendes (in preparation) and Mendes and Lissak (in preparation).

The measures of procedures and resources used in this study are as follows: Subjects were asked to indicate how satisfied they were with six military benefits or resources (pay, promotion, personnel evaluation ratings, postings, retirement benefits and resources needed to perform one's job). The participants were then asked to indicate how satisfied they were with the procedures used to allocate these six resources. The scales used were 5-point scales ranging from "Very Dissatisfied" to "Very Satisfied". General attitude was measured using the Job Descriptive Index (Smith, Kendall and Hulin, 1969).

Results

The results of the survey show that procedures are important predictors of job attitudes. Table 1 shows the results of hierarchical regression analyses predicting theta, the underlying satisfaction trait of the Job Descriptive Index, from measures of reactions to procedures and from measures of reactions to outcomes. The overall multiple correlation when all measures of both outcome and procedure are used to predict job satisfaction is .657. The hierarchical analyses test whether each set of predictors makes a significant, unique contribution to explaining the variance in job satisfaction. In the first analysis, nine measures of reactions to procedures (satisfaction with each of the six specific procedures, agreement with military procedures, an overall measure of say in determining outcomes, and an overall rating of satisfaction with treatment) were entered first in the regression equation. The resulting multiple correlation was .639. In the second step of this analysis, seven measures of reactions with outcomes (satisfaction with each of the six specific outcomes and an overall measure of satisfaction with Canadian Forces benefits) were entered in the regression equation. The increase in the multiple correlation between the first and the second step is an indication of the extent to which the predictors entered in the second step make a unique contribution to the prediction of job satisfaction. As can be seen in the table, reactions to outcomes made very little difference in the regression equation; a test of the significance of the increase in the multiple correlation fell far short of significance.

Table 1

Hierarchical Regression Predicting Work Attitudes

Predictor Set	Step	R	ΔR^2	F
Analysis 1				
Procedure	1	.639	-	-
Outcome	2	.657	.024	1.47
Analysis 2				
Outcome	1	.509	-	-
Procedure	2	.657	.173	8.29*

*$p < .01$
Analysis 1 df (7,245)
Analysis 2 df (9,245)

The second set of entries in Table 1 shows the results of a hierarchical analysis that tested the unique contribution of reactions to procedures in predicting job satisfaction. In step one, only the outcome measures were entered in the equation, with a resulting multiple correlation of .509. The addition of the procedure measures in step two of this analysis resulted in a relatively substantial increase of the multiple correlation. As reported in the Table, the test of the increase in the multiple correlation was significant.

Discussion

We began the present study with the hypothesis that reactions to procedures have a substantial effect on work attitudes. The results just presented are entirely supportive of that hypothesis. No single study can prove that reactions to organizational procedures are a major cause of job satisfaction, but the present data exceeded our initial expectations in its support of that idea. Not only do procedures appear to be equal to outcomes in their effects on work attitudes, they appear to be more important. Additional research will be needed to replicate this finding and to assess its generality, but the present data leaves little doubt that further investigation of the role of procedures is worth doing.

This study clearly establishes the importance of studying the manner in which the affective responses of members of an organization are influenced by characteristics of that organization. In making this statement it should be reiterated that no one single study can adequately address all questions. This study did not, for example, investigate questions of causality or variation in specific procedures. Nevertheless, this data points to a potentially fruitful avenue of research on the formation of individual attitudes.

References

Adams, J. S. Inequity in social exchange. In I. L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Homans, G. C. *Social behavior: Its elementary forms*. New York: Harcourt, Brace and World, 1961.

Lissak, R. and Mendes, H. Factors affecting Attrition/Retention in the Canadian Forces. (Technical Report in preparation).

Mendes, H. and Lissak, R. Canadian Forces Attrition/Retention Questionnaire Preliminary Results. (Technical Note in preparation), Canadian Forces Personnel Applied Research Unit. Toronto, Canada.

Assessing Motivational Skill Deficiencies in Military Trainees

Barbara L. McCombs
Denver Research Institute

Experience and research with the Air Force Advanced Instructional System (AIS) indicated that many students entering this computer-managed instructional (CMI) environment lack the basic conative, affective, and cognitive skills required to effectively motivate themselves and perform well in their technical training courses. Prior work in this area has shown that substantial payoffs in reduced training time can be achieved through self-instructional student training in time management and study skills (McCombs, Dobrovolny, & Judd, 1979). This ~~earlier work, however, investigated only a~~ small set of skills essential to effective and efficient student performance in a CMI technical training environment. The present study was part of an effort to address additional student skill training areas particularly tailored to the unique conative, affective, and cognitive skill deficiencies of those students performing in the lowest quartile on course performance measures.

The basic approach taken to the identification of likely sources of student skill deficiencies in the conative, affective, and cognative domains consisted of the following steps. First, an extensive review of literature related to underachievement and skill training approaches was conducted. Second, the performance of students performing in the lowest quartile on CMI course variables of interest (completion times and test scores) was analyzed to determine whether existing AIS individual difference variables, measured in a testing battery students completed before entering their technical training course, could reliably discriminate this unsatisfactory group's performance from the remaining 75 percent of the students. Third, instructors and students in each AIS course involved in the study were interviewed to provide a more intensive look at the kinds of student characteristics that distinguished good versus poor performers. Information from these first three steps was then used to design a set of individual difference measures related to the identified conative, affective, and cognitive skill deficiencies of the poor versus good performers and to the time and score performance variables of interest. The results of validation efforts conducted with the resulting test battery are the subject of this paper, as well as the implications of these findings for future validations of the battery and for specialized skill training programs to address the identified motivational deficiencies.

Results of the Literature Review

One goal of the literature review was to identify relevant theoretical and empirical sources suggestive of skill deficiencies commonly experienced by those students performing unsatisfactorily in instructional situations. Literature was reviewed from a variety of prevalent theoretical perspectives, including attribution theory (e.g., Bar-Tal, 1978; Covington & Omelich, 1979; Halperin & Abrams, 1978; Thomas, 1979; Weiner, 1979), information processing theory (e.g., Mischel, 1979; Rogers, 1977; Sternberg, 1977, 1979; Wittrock, 1978, 1979), cognitive-behavioral orientations (e.g., Ellis, 1977; Kendall & Hollon, 1979; Mahoney, 1977; Meichenbaum & Asarow, 1979; Woolfolk & Richardson, 1978), and theories of human development (e.g., Elkind, 1978; Erikson, 1968; Maslow, 1954; Miller, 1978; White, 1966).

This review identified the importance of the following student characteristic variables for academic achievement: (a) the extent to which students have an integrated value system; (b) the extent to which students accept personal responsibility for learning; (c) students' level of self-esteem or self-acceptance; (d) students' inherent interest in

learning or intrinsic motivation; (e) students' perceptions of the locus of responsibility for their academic successes and failures; (f) students' feelings about the amount of control they have over academic outcomes; (g) students' ability to effectively and spontaneously initiate executive processes and strategies that can be applied to problem solving or reading comprehension tasks; (h) students' ability to effectively execute skills for dealing with negative affect (e.g., test anxiety) while engaging in information processing activities; (i) students' ability to cope with and adapt to task demands; (j) students' beliefs and expectations regarding learning situations and their ability to perform in these situations; (k) students' ability to cope with stressful situations through the use of assertiveness or stress management skills; (l) students' commitment to meaningful academic and personal goals; (m) students' level of intellectual, emotional, and vocational maturity; (n) students' achievement of ego identity or personality integration; and (o) the nature of students' self-verbalizations regarding themselves, their abilities, or instructional factors.

Results of Data Analyses and Interviews

The AIS performance analyses suggested that in the conative domain, compared to students performing satisfactorily, students performing unsatisfactorily had low interest and motivation toward learning the course materials. In the affective domain, the data indicated that students in the unsatisfactory versus satisfactory group experienced high anxiety toward the course and toward taking tests. In the cognitive domain, results indicated that unsatisfactory versus satisfactory groups had poor logical reasoning, reading comprehension, and study skills. In addition, a greater percentage of younger students and students with less education were in the unsatisfactory performance groups.

Both instructor and student interviews indicated that the kinds of students having the most difficulty successfully completing their technical training course were those who exhibited the following characteristics which distinguished them from students performing well. In the conative domain, the poorer students consistently were those with low motivation to learn, with few military or personal goals, and who could be classified as being low in maturity, with little self-disicpline or the ability to take responsibility for their own learning. In the affective domain, the poorer students were generally those with high levels of anxiety toward learning and taking tests, and who lacked effective skills for coping with the demands of technical training. In the cognitive domain, the poorer students were generally those with poor reasoning and comprehension skills, and/or those who lacked decision making and problem solving skills.

Approach to the Design of a Motivational Skills Battery

Based on the results of the literature review, AIS student performance analyses, and instructor and student interviews, a set of individual difference measures was selected from available measures or designed in the case where existing measures that tapped the particular student variables of concern could not be located. In general, the measures assessed students' (a) personal values and goals, (b) psychological and vocational maturity, (c) self-esteem and self-efficacy, (d) expectations about the demands of the military, technical training, and being able to take responsibility for their own learning, (e) perceptions of their ability to deal with various sources of stress, (f) ability to make responsible decisions (be assertive), (g) achievement motivation or fear of failure, (h) success/failure attributions, (i) learning-related self-verbalizations, and (j) problem solving and critical thinking skills.

A total of 140 items were selected or designed to assess the student variables of interest. These items were grouped into the following eight scales for the purpose of test administration: (1) Reasons for Joining the Military (MILREA); (2) My Skills (MYSKILL); (3) Who Am I? (WHOAMI); (4) What's Important to Me? (IMPORT); (5) My Expectations (MYEXP); (6) Critical Thinking Skills (CRITHK); (7) Things I Say To Myself (THISAY); and (8) Attitudes and Feelings About Learning (ATTSG). The MILREA scale was developed on the basis of a similar scale used in assessing Army students' reasons for joining the military. The MYSKIL, MYEXP, and ATTSG scales were developed to assess student variables not assessed in other existing measures. The WHOAMI scale consisted primarily of items selected/modified from Shostrom's (1962) Personal Orientation Inventory and Gordon's (1965) Survey of Personal Values. The CRITHK scale consisted of developed items patterned after Watson and Glaser's (1964) Critical Thinking Appraisal. The THISAY scale was composed of developed items which were based in part on the kinds of cognitions and self-verbalizations identified by Crandell and LaPointe (1979) as being related to level of psychological functioning.

The resulting battery of items was subjected to a validation process for the purpose of identifying the smallest set of items which could (a) reliably discriminate satisfactory and unsatisfactory performance groups in two AIS courses, and (b) define particular skill training needs for those students performing unsatisfactorily. The results of this validation process are described in the following section.

Validation of the Motivational Skills Battery

Administration of Measures. The eight measures described above, which required a total of between 30 and 40 minutes for students to complete, were implemented at the beginning of the routine preassessment testing procedure for students entering the Inventory Management (IM) and Precision Measuring Equipment (PME) courses. Course supervisors and instructors responsible for preassessment testing in each course volunteered to administer the measures at the beginning of preassessment for the required period of time to collect adequate student data for validation purposes. They were reminded of project goals, told of progress to date, and were informed of the purpose of the individual difference testing and procedures to be followed in collecting data on the measures. Testing packages were prepared for each course and contained a complete set of directions, the eight tests, and two AIS answer sheets. Students were instructed to complete the eight tests, in a specific order, and were told there was no time limit. Instructors explained to students that this testing was being done for a research project and their answers would be kept confidential.

Computer-based procedures for creating a separate study file of the individual difference measure data, at the item level, were developed to enable validation analyses to be conducted as part of the AIS data analysis capability. The study file was designed to be compatible with requirements SPSS (Nie et al., 1975) and the AIS Test Item Evaluation program, as well as to allow easy merging with regular AIS student performance data. During the period of data collection, data on 195 IM students and 117 PME students were collected for analysis. Performance data were available for all six blocks of the IM course. Due to the much longer length of the PME course (30 weeks) and the lower student flow, performance data were available in sufficient quantity for only the first block. Criterion variables for the predictive analyses were times-to-complete each block and block test scores.

The primary questions being addressed in this validation of the 140 items, conceptually divided into eight scales, were: (a) How reliable were the scales initially? (b) What construct validity could be demonstrated for the items? and (c) How well did the

items discriminate between students performing satisfactorily and unsatisfactorily in the IM and PME courses?

Initial Reliability of the Scales. The AIS Test Item Evaluation program was used in the calculation of means, standard deviations, and alpha reliability coefficients for each of the eight scales. In both the IM and PME courses, moderately high internal consistency (more than .75) was found for the MYSKIL, WHOAMI, IMPORT, and THISAY scales; moderate internal consistency (between .65 and .75) was found for the MILREA scale; and low internal consistency (less than .50) was found for the MYEXP, CRITHK, and ATTSG scales. These results provide some support for the conceptual classification of the items, but also point to the fact that improvements in internal consistency could be made. In addition, the similarity of findings for the two courses suggests that students were interpreting items in a similar manner.

Initial Construct Validity. The construct validity of the 140 items and their defining scales was first examined in a factor analytic process which successively compared factor structures for various combinations of scale items until consistent factors within and across courses were identified. In all factor analyses, the Varimax rotation procedure was used and only factors achieving an eigenvalue greater than 1.0 were selected. In addition, variables were considered to define a factor only if their factor loading was equal to or greater than .40. Using this procedure, a total of 30 consistent factors were identified across the two courses' successive factor analysis runs. The following number of stable factors were identified for the eight scales: (a) MILREA—5 factors; (b) MYSKIL—2 factors; (c) WHOAMI—5 factors; (d) IMPORT—4 factors; (e) MYEXP—2 factors; (f) CRITHK—2 factors; (g) THISAY—2 factors; and (h) ATTSG—1 factor. This resulting set of 30 factors left 22 items not consistently loading on these factors, or a total of 52 variables to be examined in the next step of the validation.

Predictive Validity. The basic question to be answered in the predictive validity process was whether the 52 variables identified in the initial construct validity analyses could reliably discriminate students performing staisfactorily in each course (fastest 75%, high scoring 75%) from those students performing unsatisfactorily (slowest 25%, lowest scoring 25%). For both the IM and PME courses, frequency distributions were first calculated on times-to-complete each block and block test scores, and score ranges for the satisfactory and unsatisfactory groups were defined. Discriminant analyses were then calculated, using the 52 variables from the motivational skills battery in the discrimination of satisfactory and unsatisfactory time and score groups. Next, the variables that best discriminated the groups were identified for each course. Finally, the complete set of variables that best discriminated satisfactory and unsatisfactory groups across the two courses were identified.

The discriminant analysis results across the block time and score analyses for the IM and PME courses indicated that 30 of the 52 original variables were consistently entering into the set of variables which discriminated the unsatisfactory and satisfactory student groups at the $p<.10$ level. These 30 variables encompass a majority of the characteristics which were identified in the literature as possible indicators of students in need of special skill training because of deficiencies in motivation. For the IM data, the best 30 variables correctly classified between 71.8 and 81.3 percent of the students at the $p < .001$ level. For the PME course data, the best 30 variables correctly classified between 85.9 and 89.9 percent of the students at the $p < .001$ level. Across the two courses, the set of variables which consistently discriminated student groups tapped the following areas: military expectations, self-responsibility and perceived efficacy for learning, self-esteem, motivation for self-growth or improvement, decision making skills,

positive versus negative self-verbalizations, ability to handle test anxiety, and presence or absence of academic goals.

Implications of Findings

The initial validation of the motivational skills battery indicated that a reduced set of the original items (30 variables) is predictive of the kinds of students performing satisfactorily versus unsatisfactorily in a CMI technical training environment. The variables making up the best predictor set are those theoretically related to student motivation and achievement, and they have been used following this validation effort to define a motivational skill training package for technical training students. Implementation of this package with PME students led to the findings that this type of training can significantly improve student test scores and reduce test failure rates (McCombs, 1982). What remains to be demonstrated, however, is the ability of the validated predictor set to identify those students who would most benefit from particular subsets of motivational skills training motivational deficiencies. A fruitful line of future research, therefore, is one directed at refining the revised motivational skills battery for use in individualizing student assignment to specific kinds of training addressed in the motivational skill training package (e.g., career exploration, stress management, goal setting, effective communication, problem solving). Not all students with motivational problems require training in all these areas. Individualized assignment promises to improve the efficiency of this type of training.

References

Bar-Tal, D. Attributional analysis of achievement-related behavior. Review of Educational Research, 1978, 48(2), 259-271.

Covington, M. V., & Omelich, C. L. Effort: The double-edged sword in school achievement. Journal of Educational Psychology, 1979, 71(2), 169-182.

Crandell, C. J., & LaPointe, K. A. A discrimination of cognitions: Depressive, neurotic and normal. Paper presented at the annual meeting of the American Psychological Association, New York, September 1979.

Elkind, D. Understanding the young adolescent. Adolescence, 1978, 13(49), 127-134.

Ellis, A. Rational-emotive therapy: Research data that supports the clinical and personality hypotheses of RET and other modes of cognitive-behavior therapy. The Consulting Psychologist, 1977, 7(1), 2-21.

Erikson, E. H. Identify, youth and crisis. New York: Norton, 1968.

Gordon, L. V. Survey of personal values. Chicago, IL: Science Research Associates, Inc., 1965.

Halperin, M. S., & Abrams, D. L. Sex differences in predicting final examination grades: The influence of past performance, attributions, and achievement motivation. Journal of Educational Psychology, 1978, 70(5), 763-771.

Kendall, P. C., & Hollon, S. D. (Eds.) Cognitive-behavioral interventions: Theory, research, and procedures. New York: Academic Press, 1979.

Mahoney, M. J. Reflections on the cognitive-learning trend in psychotherapy. American Pychologist, 1977, 32(1), 5-13.

Maslow, A. H. Motivation and personality. New York: Harper & Row, 1954.

McCombs, B. L. Enhancing student motivation through positive self-control strategies. Paper presented at the annual meeting of the American Psychological Association, Washington, DC, August 1982.

McCombs, B. L., Dobrovolny, J. L., & Judd, W. A. Computer-managed instruction: Development and evaluation of student skill modules to reduce training time. AFHRL-TR-79-20. Lowry AFB, CO: Technical training Division, Air Force Human Resources Laboratory, August 1979.

Meichenbaum, D., & Asarow, J. Cognitive-behavioral modification and metacognitive development: Implications for the classroom. In P. C. Kendall and S. K. Hollon (Eds.), Cognitive-behavioral interventions: Theory, research, and procedures. New York: Academic Press, 1979.

Miller, J. P., Piaget, Kohlberg, & Erikson: Developmental implications for secondary education. Adolescence, 1978, 13(50), 237-250.

Mischel, W. On the interface of cognition and personality. American Psychologist, 1979, 34(1), 740-754.

Nie, W. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. S., & Bent, D. H. Statistical package for the social sciences (2nd Edition). New York: McGraw-Hill Book Company, 1975.

Rogers, T. B. Self-reference in memory: Recognition of personality items. Journal of Research in Personality, 1977, 11(3). 295-305.

Shostrom, E. L. Personal orientation inventory. San Diego, CA: Educational and Industrial Testing Service, 1963.

Sternberg, R. J. Stalking the IQ quark. Psychology Today, 1979, 13(4), 42-54.

Sternberg, R. J. Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Lawrence Erlbaum, 1977.

Thomas, A. Learned helplessness and expectancy factors: implications for research in learning disabilities. Review of Educational Research, 1979, 49(2), 208-221.

Watson, G., & Glaser, E. M. Watson-Glaser critical thinking appraisal. New York: Harcourt, Brace & World, Inc., 1964.

Weiner, B. A theory of motivation for some classroom experiences. Journal of Educational Psychology, 1979, 71(1), 3-25.

White, R. W. Lives in progress (2nd Edition). New York: Holt, Rinehart & Winston, 1966.

Wittrock, M. C. The cognitive movement in instruction. Educational Psychologist, 1978, 13, 15-29.

Wittrock, M. C. Applications of cognitive psychology to education and training. In H. F. O'Neil, Jr. and C. D. Spielberger (Eds.), Cognitive and affective learning strategies. New York: Academic Press, 1979.

Woolfolk, R. L., & Richardson, F. C. Stress, sanity and survival. New York: Monarch, 1978.

AD P000893

ISSUES CONFRONTING THE DESIGN OF A NEW BATTLE DRILL TRAINING SYSTEM*

Larry L. Meliza, G. David Hardy,
David Johnson and Jack H. Hiller

U.S. Army Research Institute
for the Behavioral and Social Sciences

Army units organize their combat training around two programs: (a) individual soldier training based on Soldier's Manual (SM) tasks, and (b) collective training based on their Army Training and Evaluation Program (ARTEP). The ability of units to meet individual and collective training requirements is reduced by shortages of experienced trainers. peacetime garrison/administrative distractions from training, and personnel turbulence (Funk, Johnson, Batzer, Gambell, Vandecaveye and Hiller, 1980).

Effective training in the unit training environment depends on the degree to which training and evaluation can be standardized across units, and it depends on the extent to which individual training and collective training can be successfully integrated. The Chief of Staff of the Army, General Meyer, called for efforts to integrate individual and collective training in his White Paper dated February of 1980, and in a subsequent letter, dated June of 1980, called for efforts to standardize Army training.

A standardized training system would, in effect, remove much of the burden of preparing training exercises from the shoulders of inexperienced junior leaders. Such a system would also insure that soldiers entering a unit would have a training history similar to that of the unit being entered and alleviate many of the training problems caused by personnel turbulence. Further, a standardized training system would reduce the amount of time required to plan/prepare effective training exercises and help to compensate for garrison/administrative requirements which disrupt training schedules and reduce the time available for training. Integration of individual and collective training could insure that soldiers have mastered those individual skills necessary to benefit from collective training and even make it possible for training on selected individual tasks to be conducted concurrently with collective training.

The goal of the present project was to develop a standardized training system which integrates both individual and collective training requirements in small units (e.g., squad, armor platoon, section, crew). The focus of the effort was collective training, with individual skills training subordinated to collective training requirements. The proponent for this research was the U.S. Army Training Board (ATB). ATB required a product in the form of guidance materials which training developers across U.S. Army schools could use to apply the standardized, integrated training system concept to their branches.

Inadequate or inappropriate utilization of new training innovations is a frequent and well documented problem (McCluskey and Tripp, 1975; Bialek, Brennan and Hiller, 1979; Scott, 1981). It was decided at the outset of the current

*The views expressed here do not necessarily represent those of the U.S. Army Research Institute or the Department of the Army.

project that concern over the utilization of a product should influence the early stages of product development, in that any new training system should be designed to have a high potential for utilization in the field. This decision meant that system characteristics which might be ideal from the training technician's point of view had to be compromised to mesh with the less than ideal circumstances in which unit training exists. Designing a system which is compatible with the training environment would insure that the system is usable and is perceived by users as being a product which meets their real needs. Designing a system in this way also supports the eventual implementation of the system, since the need for implementers to first radically alter the skills/perceptions of the intended users or existing Army training management/organization (Gray, 1981) is carefully eliminated/reduced.

This project started with the design of a concept for standardizing small unit training and integrating individual skills training with collective training. The system concept was then further developed through trial application to a sample branch of the Army, light infantry. After a usable prototype system had been developed for light infantry, the principles/rules used in preparing the final system were recorded in the form of a draft guideline for training developers. The clarity/adequacy of this guideline was then tested through trial application, and necessary revisions were made in the guideline.

DESIGN OF THE SYSTEM CONCEPT

The starting point for this effort was a careful analysis of the tasks a trainer must perform to plan, prepare, and conduct integrated small unit training. The primary sources of this information were the various documents describing the Army's Battalion Training System. Given that the purpose of the project was to develop a standardized training system, the next step was to determine the extent to which these trainer tasks had been standardized or could be standardized within the framework of ARTEP documents.

After a careful review of ARTEP 7-15 for infantry units and ARTEP 71-2 for Mechanized Infantry, it was decided that increased standardization of entire ARTEP missions would not meet the need for a standarized, training-environment-compatible, small unit training system. ARTEP mission training objectives contain variable task, conditions and standards statements necessary to describe the diverse situations in which a unit must be able to perform each of its missions. If entire ARTEPS were standardized to the degree necessary to help inexperienced trainers conduct training and to reduce the effects of personnel turbulence, ARTEPS would become extremely large, cumbersome documents. Time constraints would force leaders to select among a large number of potential training objectives, and, as a result, training would not be standardized across units in terms of the specific training objectives being trained/evaluated.

It was decided to select small "chunks" of battle actions which, if standardized, would provide the greatest benefit to small unit training. Two criteria were believed to be of special importance in selecting such chunks of battle. First, the chunks of battle selected should require specific, active participation by all, or nearly all, unit members. This criterion would insure that all unit members would benefit from taking part in training. Second, the portions of battle selected should have wide applicability across ARTEP missions. In selecting these mission chunks, the small unit training vehicle would be one which fit the general rubric of "battle drills." The primary distinction between the present battle drill training system concept and battle drills informally

used by various unit leaders was in terms of the intended degree of standardization.

The selection of drills as a small unit collective training vehicle meant that the goal of integrating individual and collective training would be accomplished by integrating individual skills training with drill training. The set of SM tasks potentially covered within the small unit training system was thus reduced to those which are drill relevant. It was further determined that individual skills training could be integrated with drill training in three ways. First, certain SM tasks must be trained/evaluated in preparation for drill training to avoid tying up the collective training with individual training. Second, certain SM tasks could be completely trained/evaluated to SM standards by simply embedding them in the drill standards. Third, certain SM tasks could be fully trained/evaluated as time permits after partial coverage during drill training.

It was determined that the appropriate method of integrating a particular individual skill would depend upon identifiable characteristics of the individual skill. A decision rule was developed to determine how each soldier's manual task needed for a drill was to be integrated (i.e., as a drill prerequisite, embedding it in the drill, providing partial coverage in the drill with a recommendation to finish training as time permits). The primary goal of the decision rule was to insure that a particular individual skill would not disrupt drill training per se, or cause drill-training-time to be used in an inefficient manner.

DEVELOPMENT OF THE SYSTEM CONCEPT

Based on the definition for drill tasks formed early in the project (see Table 1), the ARTEP for light infantry squads was analyzed to identify squad/fireteam level drill candidates. Twenty-five candidates were found and then reduced to 16, with the assistance of Army Training Board subject matter experts. By retrospective analysis, the rules for identifying drills through analysis of ARTEP missions and for preparing standardized drill training objectives were developed.

TABLE I

CHARACTERISTICS OF A DRILL TASK

- Keyed to one or more ARTEP mission tasks
- Requires performance by most or all unit members
- Requires rapid unit reactions to enemy threat or leader order
- Minimizes need for leader tactical decisions and coordination with other units
- Requires a relatively standard set of actions in a variety of situations
- Has natural starting and stopping points
- Maximizes application across ARTEP Missions

The prototype drill training objectives included administrative conditions for conducting training, as well as traditional tactical conditions. The prototypes provided a brief description of desirable training site features, instructions for properly positioning the unit and the opposition force at the start of the drill, and the instructions to be given to the unit and to the opposing force. The precisely defined administrative conditions served to provide information which inexperienced junior leaders need to conduct training exercises that provide meaningful training to meet the performance standards. The prototype training objectives were reviewed by subject matter experts (SMEs) and a few minor changes were made in the content of the training objectives in response to SME feedback.

In the course of preparing prototype drill training objectives, it became apparent that certain portions of the ARTEP selected for drills were too complex to be directly covered by standardized drill training objectives. This complexity was due to the large number of different tactical situations possible. It was decided to simplify these complex training objectives to facilitate standardization and make it easier for trainers to conduct drill training by narrowing the scope of those battle chunks initially selected as candidate drills. This decision represented another compromise made to produce a usable system, since it had the effect of reducing the number of drill-relevant SM tasks and reducing the extent to which individual skills training and collective training would be integrated within the drill training system.

While defining prototype light infantry squad drills, it became apparent that relatively few individual skills could be included in drills without detracting from the objective of using drills as a collective training vehicle. A substantial number of SM tasks were excluded from drill training because including them would have required drill trainers to spend an excessive amount of time training or evaluating each individual, at the expense of collective training. Including certain other SM tasks in drills would have made it necessary for trainers to bring cumbersome equipment to the field, without supporting collective training. Other SM tasks could simply be more efficiently trained/evaluated using resources best used in garrison. Of these SM tasks found appropriate for training/ evaluation in the field, only a few could be completely covered by drill performance standards, because the SM tasks standards often require performance of actions not relevant to a given drill.

It was recognized that the act of merely placing battle drill training objectives in the hands of junior leaders was not sufficient to insure that effective drill training would be conducted. Four major potential problems in the execution of drill training were identified. First, junior leaders might lack the degree of familiarity with tactical doctrine necessary to conduct effective drill training. Second, leaders might have difficulty controlling the execution of an exercise. Third, leaders might not know how to most easily/meaningfully apply each performance standard. Fourth, management of unit training (i.e., planning, sequencing, resourcing, etc.) is complicated, and drill training is no exception. Each of these problems was addressed. Drill Trainer's Guides were prepared for each of the sixteen prototype drills. Each Trainer's Guide provides a lesson plan which includes (in addition to a training objective) references to specific drill-relevant doctrine and step-by-step instructions for conducting training on a particular drill. An abbreviated field-expedient version of each Trainer's Guide, the Trainer's Guide Outline, was prepared for use by trainers during the conduct of training. Guide Outlines were bound

together in the form of a pocket-sized booklet. An additional booklet, entitled Drill Evaluator's Checklists, was prepared for use by training supervisors to evaluate unit performance on a drill at the end of training. This latter booklet is a greatly abbreviated version of the Trainer's Guide Outline, omitting such features as the step-by-step procedures for conducting drill training. Finally, a Drill Training Management Guide was prepared to help leaders resource and schedule drill training in an efficient manner. These four training system aids combined to form a prototype Drill Training Package (DTP).

The prototype DTP was tried out within two companies of one battalion within the 7th Infantry Division. Companies were free to use or not use the DTP, at the option of leaders, during a two week period of training away from their home station. Training was observed on a non-interference basis. Both companies made extensive use of the Drill Trainer's Guide Outlines and Drill Evaluator's Checklist during the tryout. As a result of feedback provided by trainers, seven minor editorial changes were made in the content of the Guide Outlines and Checklists.

The principles/rules used in preparing the prototype DTP were recorded in the form of a draft "Guideline for Designing Drill Training Packages." The clarity/adequacy of this guideline were tested using contract staff simulating the role of school training developers. Members of the contract staff used the draft guideline to prepare sample drill training objectives for both light infantry and mechanized infantry units. Certain critical difference were found between the training objectives produced by a contract staff and the prototypes. In general, the sample training objectives were very complex and left much of the responsibility for designing drill training exercises on the shoulders of trainers. In effect, the sample training objectives were too similar to their parent ARTEP training objectives. In discussion with members of the contract staff, it became apparent that the failure to adequately specify the administrative conditions under which each drill should be conducted was due to the complexity of the sample training topics. In response to these findings, the draft guideline underwent considerable revision to explain/demonstrate the required simplicity of drill training objectives relative to ARTEP mission training objectives.

UTILIZATION OF SYSTEM CONCEPT

Soon after the company level tryout, the parent battalion and the parent brigade adopted the prototype DTP for use in training. The second resident brigade later adopted the DTP for use, as did the 1st Brigade of the 82nd Airborne Inf Division. To date, a total of over fifteen hundred copies of the DTP have been requested for use by units in the 7th Infantry, 9th Infantry, 4th Mechanized Infantry, 82nd Airborne, 101st Airborne, California National Guard, Pennsylvania National Guard and Oregon National Guard.

The U.S. Army Training and Doctrine Command (TRADOC) distributed six hundred additional copies of the DTP across major Army commands for purposes of review. Feedback received from these major commands has been highly favorable. ATB has decided to publish the revised "Guideline for Designing Drill Training Packages" as a TRADOC Pamphlet and is considering the possibility of publishing it as a Regulation. The U.S. Army Armor School has now used the guideline in preparing drill training objectives for Armor platoons.

References

Bialek, H.M., Brennan, M. and Hiller, J.H., Development and Implementation of a Performance Based Training and Evaluation System for the Combat Arms. ARI Technical Report, TR-79-B4, May 1979.

Funk, S.L., Johnson, C.A., Batzer, E., Gambell, T., Vandecaveye, G. and Hiller, J.H., Training Detractors in FORSCOM Divisions and How They are Handled, ARI Research Report 1278, May 1980.

Gray, T., Implementing Innovations: A Systems Apporach for Integrating What is known. Journal of Technology Transfer, 1981, 6. 19-32.

McCluskey, M.R., and Tripp, J.M., An Evaluation of the Utilization, Maintenance and Perceived Benefits of the Training Extension Course (TEC). Technical Report 75-18, Human Resources Research Organization, June, 1975.

Scott, T.D., Implementing Innovations in the Army: A Case Study. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, California, August, 1981.

AD P000894

A METHOD FOR IMPROVING SOLDIER'S MANUALS [1/]

Elmo E. Miller

Human Resources Research Organization

Problem

This research addresses the perennial problem of how to develop effective printed instructions. It is directed specifically at task summaries for Soldier's Manuals, but the method is applicable to virtually any kind of military task. We also wanted to translate the method into a guidebook that would be a practical help for Army writers.

Approach

Our basic approach was: (a) to revise a wide variety of task summaries, trying for radical simplicity, and (b) to formulate rules that were inherent in each revision. This is quite different from the "armchair" ruminations on which most guidance for writers is based. We also found some general principles that incorporate many of the particular rules.

Some rules and examples

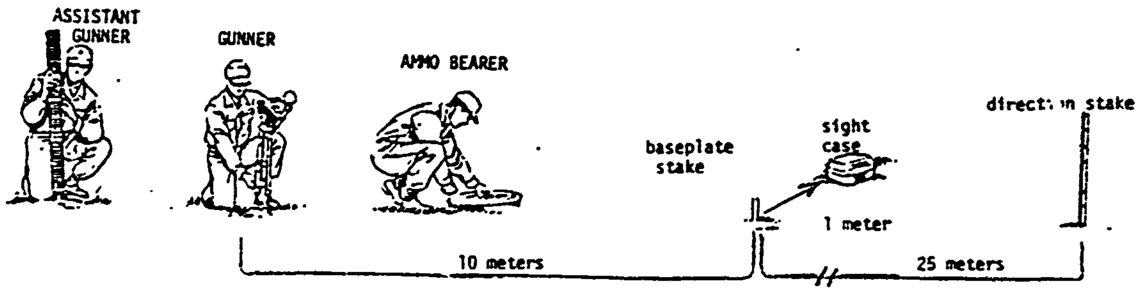
The first rule is to reduce all instructions to the barest essentials in both words and pictures. Such instructions must involve a clear path that leads the reader through the essential steps to the task objective. The reader should not be led back and forth between text and illustrations, between the text and various notes, or between alternate descriptions of the procedure. Resolve that there is no "safe" way in your basic instructions to provide extra material or alternate routes, just in case someone might need them. A writer must rely on each element of the instructions to carry the message, or find a better way to say it.

For example, the first illustration in Figure 1 combines several pictures from the original. It shows all performers and all items of equipment at the start. Since the purpose is to show configuration of these elements, the component pictures can be small, because the reader only needs enough detail to recognize the elements. Labels are provided for each person or item of equipment, so the reader is not required to go back and forth between text and illustration or between the illustration and a legend. Notice that this differs from the common practice of using call-out numbers, which are an arbitrary code requiring many extra steps.

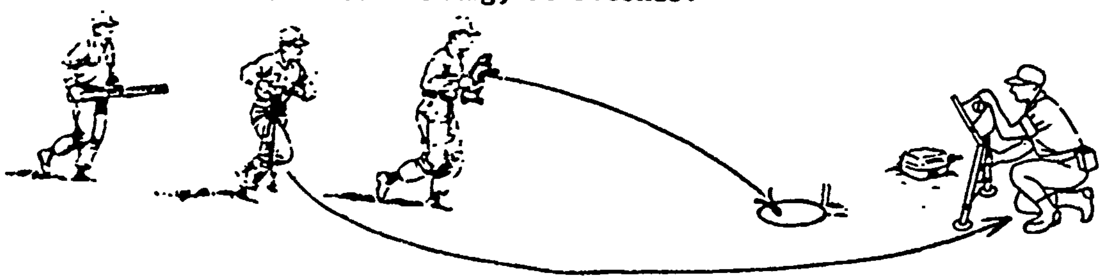
[1/] This research was conducted for the U.S. Army Research Institute for the Behavioral and Social Sciences under Contract No. MDA 903-79-C-0191 monitored by Dr. Charles O. Nystrom. The findings in this report are not to be construed as an official position of the Department of the Army unless so designated by other authorized documents.

GROUND-MOUNT 81-MM MORTAR

Conditions: You are the gunner in this situation.



2. Assembling Mortar
Squad leader: "TO YOUR FRONT." (crew stands up) "ACTION."
Scorer starts timing, 90 seconds.



•hook chain and spread bipod legs

Figure 1. Revision, Mortar Example

This also exemplifies the keystone of our method: the integration of text and illustrations. What is said with illustrations is not repeated in the text. The writer must decide in each instance which will work better, illustrations or text, and bet on that way. This may appear to be only common sense, but it is contrary to standard practice. This rule leads to tremendous simplification in most kinds of instruction. In the second illustration (of Figure 1) for example, those two arrows replace a lot of complicated prose.

Another means of simplifying is to focus on the results of each step, and leave out all trivial manipulations. For example, the revision says "hook chain" (bottom of Figure 1) instead of "Kneeling on his right knee in front of the bipod and supporting it with his left hand, he unhooks the chain, unwinds it, and rehooks the end hook on the chain hook." The result is a hooked chain, and that's all the soldier has to remember.

Note also that we used a dot as a handy way of indicating which steps will be scored. This way, the soldier knows which steps constitute the standard as he reads them.

Notice the continuity between the illustrations in Figure 1. There should be no abrupt changes in viewpoint (i.e., camera angle) between successive illustrations unless that change is clearly indicated. One picture in the original was very confusing because the viewpoint was flipped 180° without any indication of the change.

Our revisions used many fewer illustrations than the originals, but relied on them much more. The main reason for reducing the number of illustrations is <u>not</u> to save paper or artwork, but to eliminate the need to relate several pictures to each other. Inferring such relations is a tremendous burden for the reader. The number of sentences is reduced even more, for similar reasons.

Our method also involves hierarchical organization (i.e., "chunking") of the steps. An undifferentiated "laundry list" is generally hard to understand, and very difficult to remember. The more effective organization is quite compatible with other desirable features.

Figure 2 is another example, involving a comparison between original

<u>Original Instructions</u>

b. To check the firing wire (using M51 test set)

(1) Separate the firing wire conductors at both ends, and connect those at one end to the test set binding posts. Actuate test set. The indicator lamp should not flash. If it does, the firing wire has a short circuit (figure 2a)

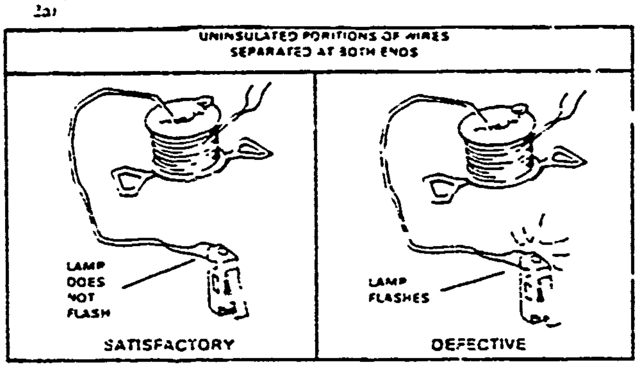


Figure 2a.

(2) Twist the wires together at one end, and connect those at the other end to the test set posts. Actuate test set. The indicator lamp should flash. If it does not flash, the firing wire has a break (figure 2b)

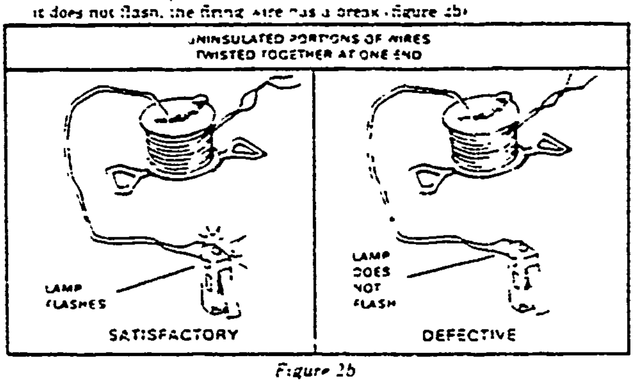
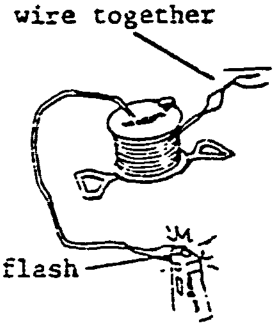


Figure 2b

<u>Revision</u>

b. To check the firing wire:

(1) Twist wire together



Test:
lamp should flash

(2) Then untwist wire and test again: lamp should not flash

Figure 2. Integration of Text and Illustration

instructions and our revision. The original has four pictures involving trivial variations, following bad advice to "use lots of pictures." There is even greater compounding of the prose, which leads the reader down an extremely devious path. The revision is considerably shorter, and the instruction "twist wires together" is connected directly to the picture with a call-out line. The revision also has a clear sequence, reading from top to bottom, without detours. This degree of simplification is not uncommon.

The same basic method applies to procedures that don't involve equipment, such as computations or making entries in a standard form. In fact, the payoff may be the greatest when no equipment is involved, because such tasks provide little intrinsic feedback, and they tend to be abstract and conceptually complex. Also, such "paperwork" tasks are often done with minimal supervision.

Figure 3 is our revision of instructions from TM38-750, on how to fill out a standard form for deferred maintenance. The most important feature is that all instructions are clustered, with each cluster connected directly to a particular response block. After each cluster there is an implied "execute" command, so the reader can respond immediately, thus minimizing the burden on memory. This differs from the common instructions to "read everything before you do anything." By sorting the information according to the reader's needs, we greatly simplify his task.

The original instructions also covered aircraft maintenance, which involves some confusing variations. But aircraft are maintained by different groups of people, so their instructions were given on a separate page.

The clusters of responses involve various kinds of subroutines, which specify behavior with exceptional precision. The first block, "nomenclature," refers to a long list of acceptable abbreviations, at the required level of generality. But this "list" kind of subroutine would not work for the date block, which requires a more generic kind of specification. Notice that these subroutines are for very familiar kinds of problems. This suggests that they have general applicability, and that a limited number of subroutines may cover most instances. Therefore, we expect to develop a highly generalizable technology as we apply our method to other tasks.

<u>A demonstration experiment</u>

Bob Cooper (who was with our organization) and I conducted a demonstration experiment to evaluate the effectiveness of this revision. Twenty-six students and professors at U.T. Austin judged the correctness of six entries on the standard form, under each of two conditions: by referring to the original instructions, and by referring to the revision. Order of presentation was balanced. We scored the number correct under each condition, and asked them to rate their confidence, and to indicate which set of instructions was easier to follow. The revision was significantly better on all three measures ($p < .01$, sign test). It reduced errors by 64%.

TASK 143: Fill Out DA Form 2408-14

Conditions: In filling out DA Form 2404 there was a fault that could not be corrected immediately, but the equipment was still operable. (This page is for equipment other than aircraft. Aircraft faults are discussed on the following page.)

Standard: Check correctness of entries on all points specified below.

Procedure:

1. Identification of vehicle

Use nomenclature and model numbers from TM 38-750, Appendix E

From vehicle inspection log, or from nomenclature plate on vehicle

2. Entries

Use one of the following:

- ⊗ = deferred maintenance on action, equipment still operable
- / = deficiency that only degrades efficiency
- — = potentially dangerous

— defined in TM 38-750 para 4-2 c(1)

Do not use "x" because that means the equipment is inoperable, and you should use DA Form 2407.

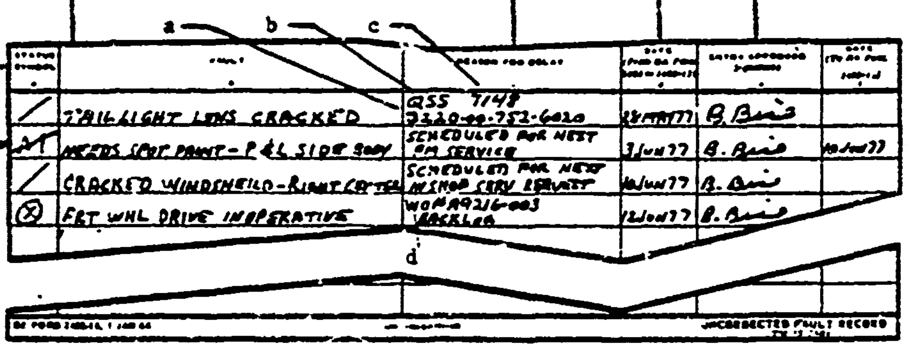
Do not erase status symbol if it is an error. Instead, draw a line through the whole entry and start again on the next line.

Copy verbatim from 2404, column c.

Date of entry

Give reason for delay and information about action taken.

Signature of commanding officer or designated representative



a. NSN (National Stock Number)

b. Julian date that part was requested

c. QSS or SSSC

d. Work order number (in case of backlog)

3. Fault Corrected

Then the person who does it writes the date in the last column and his last name initial over the status symbol.

M

10 JUN 77

4. Disposition: Six months after the last fault is corrected, this form may be discarded.

Figure 3. Revision of Standard Form for Maintenance

A classification of tasks

A taxonomy of tasks was developed so that the method could be better applied to specific tasks. A basic split is between (a) procedures with equipment, (b) procedures with data, and (c) performances that are irregular in sequence. This is because involvement of equipment and sequence of steps are important considerations in writing instructions. The finest division involves 31 categories, which is too many for discussion here. However, the following classes (with examples) may indicate the general kinds of distinctions involved: construction (construct a mortar position), assembly (ground-mount a mortar), diagnosis in maintenance (electronic troubleshooting), using numerical tables (for getting a logarithm), and identification of equipment (combat vehicle identification).

Discussion

Revision is a craft involving numerous rules and principles. However, it is not some vague form of art, in which the practices are merely a matter of opinion. Revision becomes much easier as numerous examples and more explicit rules are developed. Even today there are many high-density tasks, in which people are especially dependent upon printed instructions, where this kind of revision will be well worth the effort.

The keystone of our method is integration of text and illustrations. This allows us to slip through the horns of an old instructional dilemma: whether to present rules first, or examples. Our method leads the reader to consider rules and examples together, which appears to be the best way by far.

The method may alleviate critical manpower requirements. Some of the most stringent requirements seem to result from tasks that have not been sufficiently proceduralized. It generally requires more experience and ability to develop effective procedures and to communicate them, than to perform the procedures once they are established. Our methods may be an important tool in proceduralizing of tasks.

AD P000895

RETENTION OF ARMOR PROCEDURES: A STRUCTURAL ANALYSIS

John E. Morrison

ARI Field Unit - Fort Knox

Over the last few years the nature of armor tasks has changed rather dramatically. In older tanks, tasks such as ranging to the target and leading a moving target have a large skill component. With the advent of the laser rangefinder and automatic lead components built into modern fire control systems, these tasks have become largely automated and thus easier to execute. However, the pre- and post-operation procedures required by these sophisticated systems are quite complex and difficult to learn. Complicating this training problem is the fact that procedural skills are particularly susceptible to forgetting over periods of no practice. Because of the importance of procedural skills to armor performance, the ARI Field Unit at Fort Knox has been involved in developing methods for training and sustaining procedural skills.

As a basis for this research program, Morrison and Goldberg (1982) presented a model of the memory structure which underlies procedural task performance. The model assumed that memory for a procedure is hierarchically organized around task goals. In the present study, this model was tested by a proximity analysis of soldiers' recall. Proximity analysis (Friendly, 1979) is based on the assumption that items grouped together in memory tend to cluster together at recall. To perform this analysis, estimates of temporal or ordinal proximity are obtained on an item-by-item basis. The proximities are then subjected to a hierarchical cluster analysis, the result being a graphical representation of memory structure. This technique was applied to the verbal recall and hands-on performance of armor procedures. It was predicted that soldier responses would cluster about discernible task goals.

A significant characteristic of procedural skills is their tendency to be forgotten over time. For instance, Osborn, Campbell, and Harris (1979) documented declines in armor task performance over the period between basic training and field unit assignment. Perhaps such decrements in skill are associated with changes to memory organization. To investigate this possibility, memory structures produced by armor crewmen in the final phase of entry-level training were compared to structures of armor crewmen assigned to an operational field unit.

METHOD

Testing Procedure

Two groups of armor crewmen participated in the present research project. One group was made up of 12 soldiers from the 1st Armor One Station Unit Training Brigade at Fort Knox (OSUT soldiers). The second group consisted of 12 soldiers drawn from the 194th Armored Brigade, a Forces Command unit at Fort Knox (UNIT soldiers).

Soldiers were tested on six procedures in all, but results from only two were reported here. (Results from all six tasks were presented in Morrison, 1982.) The representative tasks were to clear the M240 coaxial machine gun and to put the AN/VRC-64 tactical FM radio into operation. Soldiers were first asked individually to recall the procedures in a step-by-step manner while a tester recorded their responses on audio tape. Then, they were given hands-on tests on the same tasks. Hands-on performance was videotaped by another tester. Later, the audio and video tapes were transcribed into written protocols.

Proximity Analysis

According to Friendly's (1979) technique, proximity can be measured in terms of the differences in ordinal positions of recalled items or in terms of inter-response times. The choice of measures depended on the sequence demands of the task.

The elements of the clear task had to be performed in a fixed order, and, for the most part, soldiers recalled the procedural elements in that sequence. Consequently, adjacent elements in the protocols all had a proximity of one with respect to output order. In contrast, the time intervals between protocol elements were free to vary between subjects. For the clear task, then, proximities were defined in terms of inter-response times. However, inter-response proximities could be obtained for verbal recall and not for hands-on performance. Two problems prevented measurement of times between hands-on responses. First, the onset and offset of a response element could not be reliably observed within the fluid series of actions which comprise hands-on performance. Second, factors other than memory organization (e.g., spatial location of parts) affected inter-response times. Thus, the memory structure for the clear task was derived from verbal recall and not hands-on performance.

In contrast to the clear task, elements of the radio operation procedure could correctly be performed in various orders. Consequently, both inter-response times and output order could have been used as measures of proximity. However, output order had two advantages over inter-response times under these circumstances. First, output order was a more stable measure than inter-response time, especially without restrictions on response order. Second, output order could be measured for hands-on performance as well as verbal recall, allowing comparisons of memory structures derived from both modes of performance. Thus, output order was used for this task to derive two memory structures based on verbal and hands-on performance.

Proximities for every pair of elements were computed by taking the median of the inter-response times (clear task) or the mean of the differences in output order (radio operation task). Medians were used in the clear task because of the marked positive skew of the inter-response times. The central tendencies of the soldier proximities were then entered into element-by-element proximity matrices. A hierarchical cluster analysis was then applied to these data. The order of elements for the clear structure (left-to-right) was simply the prescribed sequence for the clear task. For the radio operation procedure, however, the displayed sequence was determined by the transition probabilities generated by the soldiers' performance.

RESULTS AND DISCUSSION

Table 1 contrasts OSUT and UNIT groups on the mean number of total

Table 1
Mean Errors in Response

Tasks	Group OSUT	UNIT	p
	Verbal Recall		
Clear the M240	1.4	3.2	<.01
Operate the AN/VRC-64	0.8	6.0	<.001
	(Hands-On Performance)		
Clear the M240	0.6	1.4	N.S.
Operate the AN/VRC-64	1.0	3.6	<.01

errors committed while either recalling or performing the procedures. As can be seen, UNIT soldiers made more errors than OSUT soldiers on every task. T-tests revealed these differences to be significant except for the contrast of hands-on performance on the clear task. These results provided further evidence that procedural skill performance does decline over the period from entry-level training to field unit assignment. Furthermore, the group differences in accuracy of verbal recall parallel the differences in hands-on proficiency.

The hierarchical structures derived from verbal recall of the clear task are shown in Figure 1. Both OSUT and UNIT structures indicate that task elements are organized around discernible, temporal subgoals. It can be seen that both structures are segmented into two high-level sequential subgoals. Elements of the first group relate to the removal of all sources of ammunition from the weapon. The second group of elements pertains to returning the weapon to a safe state after unloading. As can be seen, some of the intermediate hierarchical connections differ between OSUT and UNIT structures, but the lowest level relations show exactly the same pairings of elements. These first-order relationships reflect a few mechanical and safety rules which serve as basic constraints to order: (a) The safety must be in FIRE in order to move the bolt either forward or backward; (b) to prevent accidental discharge, the safety must be in SAFE before opening the cover; and (c) the firing chamber is accessed by lifting the feed tray.

The OSUT and UNIT structures for the radio operation procedure are shown in Figure 2. In contrast to the temporal organization of the clear task, recalled elements of the radio operation procedure are organized around the spatial relationships between the AN/VRC-64 components. In both OSUT and UNIT structures, there are three discernible subgoals which relate to major radio components: connect/adjust the audio accessories, operate the audio frequency amplifier, and operate the radio-transmitter. The latter two subgoals are joined at a superordinate level presumedly because the audio frequency amplifier is located on top of the radio-transmitter, both of which are separated in space from the crewman's control box and audio accessories. Even at the lowest hierarchical level, spatial organization is still obvious. For instance, the elements "adjust volume" and "set function switch on SQUELCH" do not have to be performed in any particular order. However, because the volume control and function switch are located close together on the radio

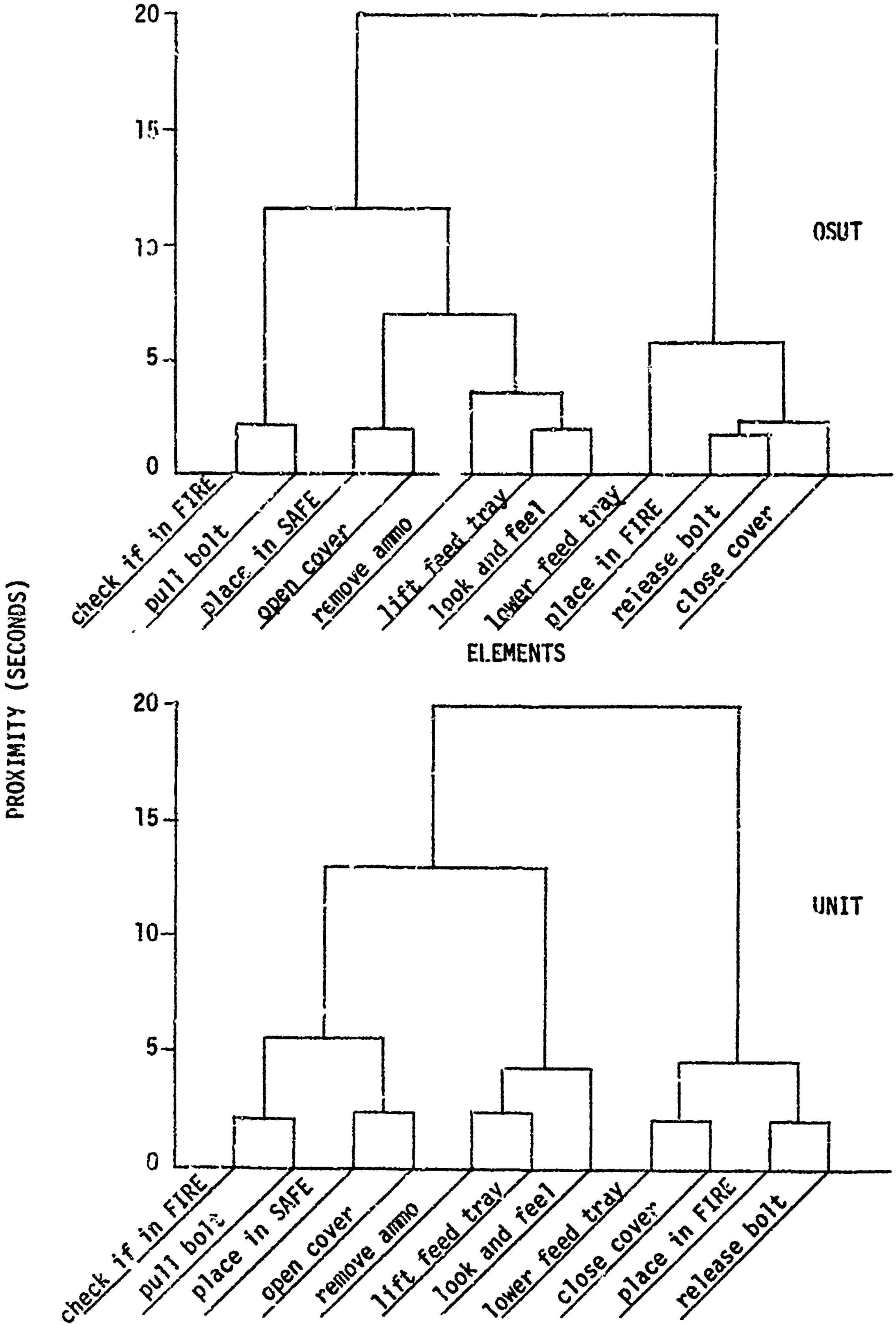


Figure 1. Hierarchical structure for verbal recall of the clearing task.

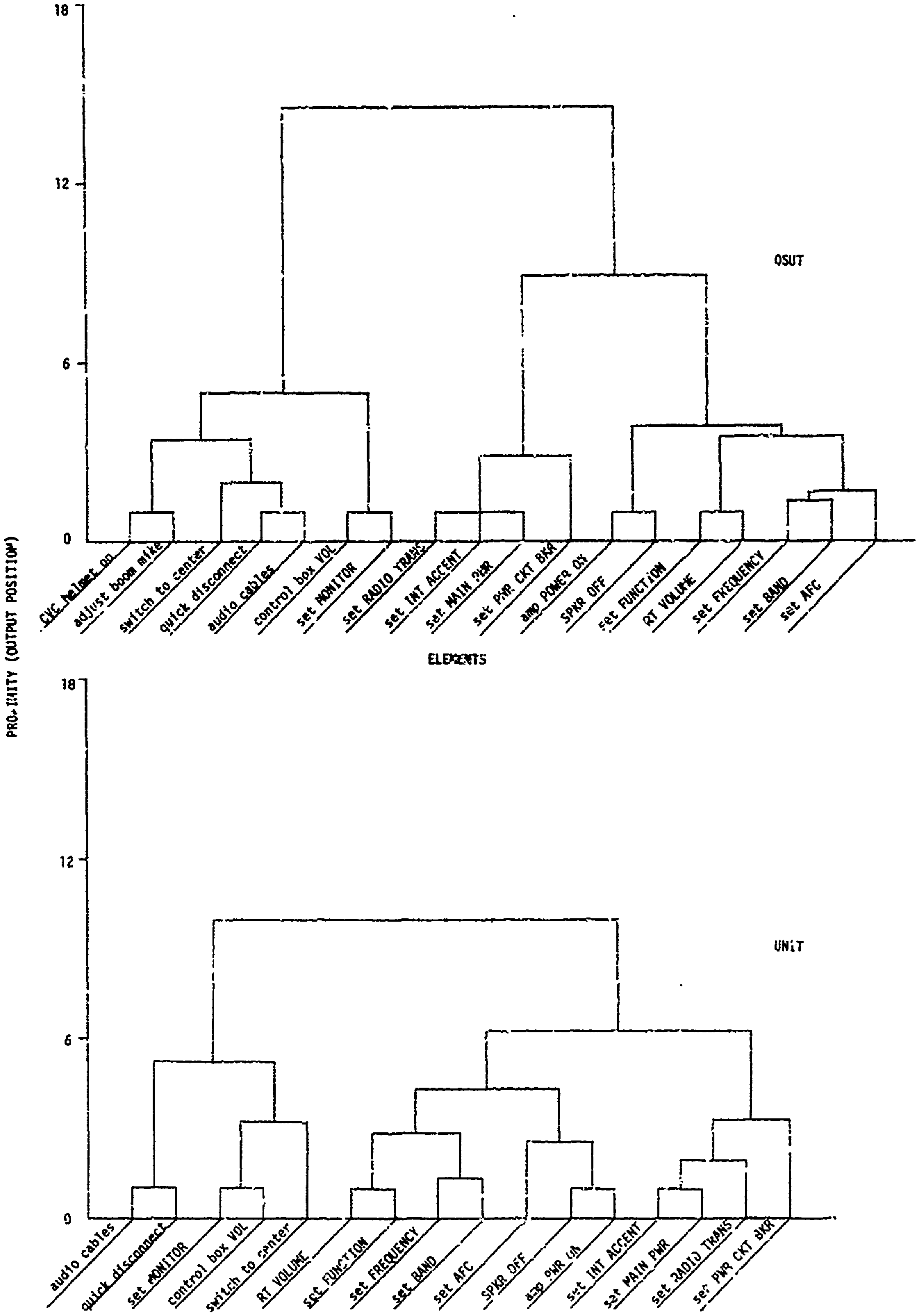


Figure 2. Hierarchical structure for verbal recall of the radio operation task.

transmitter, both OSUT and UNIT soldiers recalled the two steps together in their protocols. Consequently, these elements are directly connected at a low hierarchical level.

Although there were some minor discrepancies between OSUT and UNIT structures, the similarities between group hierarchies were more striking than the differences. To obtain a measure of structural isomorphism, entries in OSUT proximity matrices were correlated with corresponding entries in UNIT matrices. For the clear task and the radio operation procedure, the correlations were quite high (.93 and .82, respectively) indicating similar patterns of response proximities. These findings suggested that changes in recall levels do not necessarily imply changes in memory organization.

Using output order as a proximity measure, hierarchical structures of the radio task were also derived from hands-on performance. All in all, the hands-on structures were remarkably similar to the verbal structures. However, the correspondence appeared stronger for the UNIT than the OSUT soldiers. To test this, verbal and hands-on matrices were correlated for OSUT and UNIT data separately. The correlation coefficients were .95 and .75, respectively. The significance of the difference between correlations was tested by using the "jack-knife" procedure for estimating the sampling distribution. The difference was highly significant, $\underline{t}$ (14) = 23.70, $\underline{p}$ <.001. The analyses thus confirmed a high degree of similarity between verbal and hands-on structures for OSUT soldiers but a lesser degree of correspondence in the UNIT structures.

Research has indicated that making learners aware of task structure increases response organization and improves recall. Thus, structural information garnered from proximity analyses may be used to aid in training and sustaining procedural skills. However, to apply this information to a real-world training situation, task goal structures must be presented in a way that is comprehensible to trainers and students with a minimum of explanation. Future research will be addressed to designing structural training aids and determining how such aids can best be incorporated into procedural training.

REFERENCES

Friendly, M. Methods for finding graphic representations of associative memory structures. In C. R. Puff (Ed.), Memory organization and structure. New York, NY: Academic Press, 1979.

Morrison, J. E. Retention of armor procedures: A structural analysis (Draft Report). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, April 1982.

Morrison, J. E., & Goldberg, S. L. Toward a cognitive approach to armor procedural training (Draft Report). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, March 1982.

Osborn, W. C., Campbell, C. H., & Harris, J. H. The retention of tank crewman skills (Research Rep. 1234). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, December 1979.

The Relationship between Farming Activities, Physical Fitness and Body Fat

John B. O'Leary, MD
Department of Family Practice and Community Health
University of Minnesota

INTRODUCTION

The ultimate goal of this research is to develop a simple, accurate method for measuring physical fitness in large population groups.

The objectives of the present study represent small steps toward this larger goal:

- To calculate mean fitness levels of groups of National Guardsmen and to determine the statistical significance, if any, of observed differences.
- To explore the generalizability of studies on National Guardsmen by comparing height/weight data on National Guardsmen with data from the most recent probability sample of U.S. males, the 1971-1974 U.S. HANES survey (Abraham, Johnson & Najjar, 1971-1974).
- To identify military units with higher percentages of farmers and to begin pilot studies of life habits, fitness levels and body fat of men in these units. and
- To record baseline data for future calculations of the reliability and validity of this new fitness testing method.

BACKGROUND

The health status of farmers is difficult to describe and much harder to measure. Ten years ago, a group at the University of Minnesota attempted to correlate a widely used health status questionnaire with three direct measurements of physical fitness and oral hygiene in a group of farmers (O'Leary, Zaki & Alexander, 1973). The questionnaire consisted of four questions covering days of hospitalization, the history of the use of medicines, a checklist of acute conditions and a checklist of chronic conditions. Each condition was assigned a numerical weighting on the seemingly logical assumption that higher scores would represent poorer health (Kisch, Kovner, Harris & Kline, 1969). Rank order correlation of all variables revealed no significant relationship between the health status questionnaire, physical fitness measured by bicycle ergometry, oral debris as measured by staining and periodontitis evaluated by direct inspection. The failure of a standard health questionnaire to correlate with direct health status measurements suggested that farmers as an occupational group are not an homogenous subset. This inference was supported by a statement by the director of the southwestern Minnesota Agricultural Experiment Station who estimated that about half of the farmers in that area grew only crops while the other half were diversified with farm animals, usually hogs or cattle, in addition to crops of corn, soybeans, or small grain.

There are obvious differences in the appearance of these two types of farmers' land and in their work habits. Cash cropping is closely associated with flatland and a life-style which includes working intensively a few months of the year. Those farms with both crops and farm animals appear quite different from those with only crops. The more diversified farming is often associated with rolling hills, younger farmers and a work schedule that goes on for seven days a week, twelve months a year.

The medical literature of recent years is filled with reports of the relationship between work activities, leisure activity and coronary disease (Chave, Morris & Moss, 1978; Brand, Paffenbarger, Sholtz & Kampert, 1979; Magnus, Matroos & Strakee, 1979). For the last 40 years medical literature has emphasized life-style and disease rather than life-style and health. More recently, especially within the past years, many medical writers have shifted from a disease orientation. There appears to be increasing emphasis on direct measurement of health with several investigators relating life habits to measurements of physical fitness levels and body fat (DeBacker, Kornitzer, Sobolski, Dramaiz, Degre, DeMarneffe & Denolin, 1981; Leon, Jacobs, DeBacker & Taylor, 1981).

The need for accuracy in fitness testing has confined more testing to exercise physiology laboratories because no simple yet accurate technique has been available for field fitness testing. Fifty years ago, Hunt (1921) described the search for a technique of making physical fitness measurements in the field by saying that what we are really trying to measure is the capacity for moving one's body from place to place by one's customary means.

Customary activities imply motivation and a major difficulty with field tests of walking or running is the evaluation of motivation. Performance on a performance-based field test can be directly proportional to the amount of effort expended by the subject. Treadmill testing can provide an evaluation of motivation by electrocardiographic recording of maximum heart rates. Motivation can be evaluated even more accurately during maximal treadmill testing by determining the amount of oxygen an individual is able to consume when worked to exhaustion. When the maximal amount of oxygen utilized remains the same on repeated testing, it can be assumed that the individual has made a maximal effort. The amount of oxygen consumed has been shown to be the single best predictor of an individual's actual performance in exercise situations. This, then, is the gold standard . . . VO_2 max . . . the "truth" against which other fitness testing methods are judged.

The most widely used field physical fitness tests were recently evaluated by Edmond Burke (1976) who reported that among all tests, the 12-minute run correlated best with VO_2 max. Other studies have supported this and in general have concluded that field tests which correlate best with VO_2 max are tests lasting 12 to 15 minutes and covering distances of 1 1/2 to 2 1/2 miles (Cooper, 1968; Cooper & Askewin, 1966; Doolittle & Bigbee, 1968).

But all these field testing methods suffered from the same two problems, evaluating motivation and controlling pace. About 15 years ago, Sedgewick and Paddick (1966) were the first to suggest a field method for accurately controlling pace. Their pacing device consisted of a 20-foot rotating arm which paced individuals running on circular tracks. The advantages of this method were that it did involve measuring the capacity for moving one's body about by one's customary means, namely walking and running on a level surface.

It also had an advantage over laboratory tests in that it could be given in a minimal period of time to a large number of subjects. The disadvantage was the technical difficulty of constructing and controlling a 20-foot rotating arm.

During the past two years, I have developed a field method for fitness testing which replaces the 20-foot rotating arm by a timing device placed in the center of concentric circles. This timer emits a short beep every six seconds and can be calibrated so that when one crosses a diameter of a circle at six-second intervals, one is walking at a controlled speed regulated by the timer and the circumference of the circle.

This new field test works in a manner similar to the device suggested by Sedgewick and Paddick (1966). It simulates a treadmill by having subjects walk in progressively larger concentric circles Work loads can be controlled accurately and can be increased from three miles per hour to ten miles per hour by one MPH increments, four minutes at each stage. The endpoint is the number of minutes completed before subjects become too exhausted to continue.

This new method was first used to test 1,300 National Guardsmen attending annual training at Camp Ripley, Minnesota, during the summer of 1981. Approximately 1,000 of these men were from rural areas. The fitness levels in men age 18-24 were excellent but there was a sharp decline of fitness with age. As fitness levels declined, measurements of body fat by a standard skinfold method showed a marked increase in the mean percentage of body fat (Sedgwick & Paddick, 1966; O'Leary, 1982). It was these findings which raised questions as to whether rural groups would differ from a probability sample of the U.S. population in height and weight and, if so, would farm life-style account for these differences?

METHODS

Military units to be tested were selected from among southwestern Minnesota communities which were estimated by senior National Guard officers to contain the highest percentage of men actively engaged in farming. Guardsmen in these communities train on weekends. Previous experience had suggested that the circular testing pattern could be marked on the local armory floor with colored masking tape so that floors would not be damaged.

Informed consent was obtained from all subjects. Data was recorded on a pre-coded data form which included Social Security number and age. Height and weight were measured and a smoking history obtained. Farm life-style was determined by asking those who worked on farms to record how many acres they had in cultivation, how many feeder cattle, how many cow-calf, how many dairy cattle and how many hogs were on their farms?

Body fat was estimated by two methods. First method was girth measurement using standard military technique which involves duplicate measurement of the neck and waist circumference. The second method of estimating body fat involved duplicate measurement of chest, abdomen, and thigh skinfold using a Lange caliper calibrated at 7.5 grams per square millimeter of faceplate. Physical fitness levels were determined by the method previously described (O'Leary, 1982) with the endpoint being the number of minutes completed before exhaustion.

RESULTS

The 1981 study of National Guardsmen (O'Leary, 1982) suggested minor differences between the nine National Guard units (n = 464) with the highest percentage of participation in the study. However, when homogenous subsets for the variables weight and fitness were compared by analysis of variance, these differences were not found to be statistically significant. Most differences could be accounted for by age (Jackson & Pollock, 1978), with those units composed of younger individuals scoring significantly better on fitness tests and having significantly less body fat.

TABLE 1

GROUP NUMBER	1	2	3	4
Military unit of U.S. Army National Guard	682 ENG BN	1742 TRANS Co	200 ADA Arty	47 HQ DIV Arty
Number of Guardsmen Tested	46	113	36	34
Mean Age of Guardsmen in Years	30	30	36	37
Mean Height in Inches of Test Group	69.5 SD 4.5	70.3 SD 2.8	68.0 SD 2.4	70.5 SD 2.5
Mean Weight in Pounds for Test Group with Correlation for Clothing Factors	173.8 SD 28.7	181.2 SD 24.4	186.1 SD 22.5	192.0 SD 29.2
95% Confidence Interval for Mean Body Weight	165.2 to 182.3	176.6 to 185.7	178.5 to 193.8	182.0 to 202.1
Mean Body Weight for Men of Same Age and Height as Test Group from U.S. HANES 1971-1974 Survey (Abraham, et al.)	175.3 SD 23.3	180.4 SD 24.2	181.3 SD 23.5	191.6 SD 26.2
Mean Fitness Levels Recorded as Number of Minutes Completed in Incremental Walk/Run to Exhaustion, and Standard Deviation	19.2 MIN SD 4.9 MIN	18.3 MIN SD 5.0 MIN	17.4 MIN SD 5.3 MIN	17.9 MIN SD 4.1 MIN
95% Confidence Interval for Mean Fitness Level	17.8 to 20.7 Minutes	17.4 to 19.2 Minutes	17.8 to 20.7 Minutes	16.8 to 18.9 Minutes

As shown in Table 1, which compares groups of similar ages, there were no significant differences in fitness levels.

Of the 347 individuals tested in March 1982 at armories located in seven south-western Minnesota communities, there were 45 who were actively engaged in agriculture (see Table 2). The difference in age between farmers who produce cash crops and those who work with farm animals is significant.

TABLE 2

	N	$\bar{X}$ Age SD	$\bar{X}$ % BODY FAT* SD
Individuals Who Work Only with Cash Crops	10	27.7 Years SD 5.2	19.2% SD 4.3
Individuals Who Work with Both Crops and Farm Animals	35	23.0 Years SD 4.8	13.4% SD 3.9

*When Mean Body Density (MBD) = 1.10938 - 0.0008267 (X_2) + 0.0000016 $(X_2)^2$ - 0.0002574 (X_3) when X_2 = Sum of Chest, Abdomen and Thigh Skinfolds and X_3 = Age in Years. Percentage Body Fat = (4.57/MBD - 4.142) x 100 (Jackson & Pollock, 1978)

DISCUSSION

The finding that only 45 [approximately 13 percent] of men in seven rural Minnesota National Guard units were actively working on farms was surprisingly low. This may simply represent a sampling error in that many farmers may not join the National Guard. Large farm operations have such a great capital investment that there may not be a financial inducement to join a National Guard unit.

Another explanation of the low percentage of farmers might be that the massive migration from farms to cities in the United States has depleted the Minnesota farm population. Shover (1976) reported that movement from farms to urban areas has been very rapid. Outmigration reached a peak of 1 million per year in the 1950's. He states that at present, less than 5 percent of the population of the United States is employed in farming.

Vogel and Patton (1978) reported significant differences in fitness levels and body fat among regular army units. They speculated that these observed differences were related to stages of training. The insignificant differences in fitness levels noted in the present study probably reflects the fact that National Guard units tend to be at similar stages of training.

National Guard units, because they are composed of citizen soldiers, may also be more representative of the general population than other military groups. This is suggested by the nearly identical weights of the National Guardsmen and the U.S. HANES probability sample. Only Group 3, the 200th Air Defense Artillery Unit from New Mexico, differed slightly from the HANES sample. This probably reflects a basic anthropological difference between this primarily Mexican-American group and the HANES sample.

The small number of National Guardsmen actively engaged in agricultural production and the selection bias inherent in a military study limit the inferences that can be drawn from this study. The seemingly large differences in body fat between cash croppers [13.4 percent] and those working with farm animals [19.2 percent] is not signficant because body fat increases with age and the mean age of those working with farm animals was 23 years while the mean age of the cash croppers was 27.7 years [Table 2].

REFERENCES

Abraham, S., Johnson, C.L., & Najjar, M.F. Weight by height and age for adults 18-74 years. U.S. National Center for Health Statistics, Vital and Health Statistics, 1971-1974, Series 11, No. 208.

Brand, R.J., Paffenbarger, R.S., Jr., Sholtz, R.I., & Kampert, J.B. Work activity and fatal heart attack studied by multiple logistic risk analysis. American Journal of Epidemiology, 1979, 110(1):52-62.

Burke, E.J. Validity of selected laboratory field tests of physical working capacity. The Research Quarterly, 1976, 47:95-104.

Chave, S.P.W., Morris, J.N., & Moss, S. Vigorous exercise in leisure time and the death rate: A study of male civil servants. Journal of Epidemiology and Community Health, 1978, 32:239-243.

Cooper, K.H. A means of assessing maximal oxygen intake. Journal of the American Medical Association, 1968, 203:203.

Cooper, K.H., & Askewin, R. Reliability of the 600 yard run-walk at the secondary school level. Research Quarterly: American Association for Health, Physical Education and Recreation, 1966, 37:451.

DeBacker, G., Kornitzer, M., Sobolski, J., Dramaiz, M., Degre, S., DeMarneffe, M., & Denolin H. Physical activity and physical fitness levels of Belgian males age 40-55 years. Cardiology, 1981, 67:110-128.

Doolittle, T.L., & Bigbee, R. The twelve-minute run-walk: A test of cardiorespiratory fitness of adolescent boys. The Research Quarterly, 1968, 39:493.

Hunt, G.H., & Pembrey, M.S. Tests for physical efficiency. Guys Hospital Report, 1921, 71:415-429.

Jackson, A.S., & Pollock, M.L. Generalized equations for predicting body density of men. British Journal of Nutrition, 1978, 40:497-504.

Kisch, A.I., Kovner, J.W., Harris, L.J., & Kline, G. New proxy measure for health status. Health Services Research, 1969, 4:223.

Leon, A.S., Jacobs, D.R., DeBacker, G., & Taylor, H.L. Relationship of physical characteristics and life habits to treadmill exercise capacity. American Journal of Epidemiology, 1981, 113:653-659.

Magnus, K., Matroos, A., & Strakee, J. Walking, cycling or gardening with or without seasonal interruption, in relation to acute coronary events. American Journal of Epidemiology, 1979, 110(6):724-733.

O'Leary, J.B. A new pacing technique for controlling work load in running. Manuscript submitted for publication, 1982.

O'Leary, J.B., Zaki, H.A., & Alexander, J.F. The relationship between Kisch's health status proxy and three direct measurements of health status. Minnesota Medicine, 1973, 56:82-86.

Sedgwick, A.W., & Paddick, R.J. A technique for controlling work load in running. The Research Quarterly, 1966, 37(3):446-448.

Shover, J.L. First majority - last minority. DeKalb, Illinois: Northern Illinois University Press, 1976, pp 5, 11.

Vogel, J.A., & Patton, J.F. Evaluation of fitness in US Army. NATO Symposium of Physical Fitness, DS/DR 98, 29-35, 1978.

ABSTRACT

The Effect of Fatigue on Reliability of Job Inventory Responses

John J. Pass, PhD
Robert E. Chatfield

Navy Occupational Development and Analysis Center

Marjorie Royle

Navy Personnel Research and Development Center

It has long been hypothesized that the length of a survey instrument such as a Job Task Inventory or a so-called Training Importance Survey could well have an effect on the reliability of the responses on that instrument. In designing the Navy's current Training Importance Survey (TIS), a decision was made to examine this hypothesis. The approach was to administer two versions of the instrument, one containing sections of items in the reverse order of the other, and then determine if responses to items were a function of the item order of presentation. This determination consisted of comparing vectors of mean responses to items within and across forward and reverse versions and examining the reliability of items as a function of order. Findings for one particularly long instrument revealed that there was indeed a greater difference in mean vectors across as compared to within versions and that reliability in terms of number of responses made to an item was a function of order. Additional studies are ongoing to determine the generality of these findings and to determine procedures to minimize this unreliability.

A COMPARISON OF STRATEGIES FOR IMPROVING STRESS-COPING RESPONSES

Earl H. Potter III
Department of Economics and Management
U. S. Coast Guard Academy

Implicit in the many studies of the relationship between stress and performance is the belief that a better understanding of stress will support the development of treatments for stress. Foremost, these studies demonstrate that excessive stress does have negative consequences for performance. In addition, such studies have begun to detail the process whereby stress impacts on performance. In the first case we have a call for action; in the second, the tentative outline for an action plan.

Stress, according to Selye (1974), is the response of the body to any perceived demand. McGrath (1976) argues that the perception of demand is a complex process that includes the individual's estimate of the degree of difficulty of the demand contrasted with the availability of resources to meet the demand. Some situations are likely to be perceived as stressful by anyone, for example, running out of gas in the Bronx after midnight or receiving a registered letter from the IRS. The experience of stress in other situations, however, will vary widely with individuals. Meeting new people or taking an exam are common situations which terrify some while boring others.

When individuals encounter demands which they experience as stressful, their performance of a wide range of tasks tends to suffer. Decrements in performance are most severe when stress is high and the task is new and unfamiliar. Yet even in the most demanding situations some people tend to succeed. A growing body of evidence suggests that such people may actually experience a level of stress comparable to other persons who fail in the same situation. One reason for their success appears to be in their ability to focus on the task at hand and avoid distracting thoughts and behaviors directed at coping with the experience of stress itself (Anderson, 1976; Sarason, 1979). The possibility that such skills for dealing with stress might be taught to less skilled persons is suggested by the several existing approaches to dealing with test anxiety (e.g., Meichenbaum, 1972).

Other researchers have noted that seemingly exogenous influences such as available social supports can diminish the negative consequences of stress. Available social supports have been reported to minimize the impact of stress on psychological and physical health (Bloom, 1975) and to improve academic and professional performance (Goper'und, 1980). In combination, these studies suggest that a member of a cohesive team who avoids catastrophising about the consequences of failure and concentrates on meeting the challenge is the person most likely to survive stress and prosper.

This paper summarizes one aspect of a set of three studies intended to develop coping skills and social supports among the cadet corps of the U. S. Coast Guard Academy. While smaller, and popularly considered less military,

the Coast Guard Academy functions much the same as the larger Academies operating under the Department of Defense. Cadets enter the Academy at the beginning of the summer preceding their freshman year. This summer training period which precedes the start of the academic year is called "Swab Summer" and is intended, among other things, to be stressful. It has changed little since Dorbusch (1955) used the Coast Guard Academy as a model for military socialization. Swab Summer is a period of radical change, tremendous pressure for success, heavy demands on time and physical challenge. Previous studies have shown that many of the stressful elements of the summer training experience exist throughout a cadet's four years at the Coast Guard Academy. These studies have also shown that such stress is associated with a decrease in academic performance (Barnes, Potter & Fiedler, 1982). Recognition of these results presented the Coast Guard with a dilemma. On the one hand, it is desirable to maximize cadet performance, on the other hand, the Coast Guard has no intention of modifying its training program in order to reduce stress. Therefore, efforts to improve cadet performance have been focused on improving the capability of cadets to deal with existing levels of stress.

Study I was planned as a field experiment in which one treatment group would receive a manipulation intended to focus a cadet's attention on his own ability to meet and deal successfully with the challenge of Swab Summer. A second treatment group received an additional manipulation intended to increase support networks within his platoon. It was hypothesized that subjects in the treatment groups would see the Swab Summer experience as being more manageable and less stressful and report themselves to be more ready to meet the challenges of the fall semester than a control group. Furthermore, it was hypothesized that the support network treatment would result in an increase in the quality and number of social supports. Assuming that treatments were effective in changing cadet responses to stress, it was hypothesized that cadets in the treatment groups would perform better academically than cadets in the control group during the fall semester.

Study II was intended as a replication of Study I. Study III was designed to follow up on the unexpected outcomes of Studies I and II. In Study III the nature of the coping skills treatment was reversed in order to assess the impact of a different instructional set upon perceptions of stress and performance.

STUDY I
Method

Subjects. Subjects in Study I were 345 cadets of the Class of 1984 who entered the Coast Guard Academy in July of 1980. One platoon of 40 persons was omitted from the study because they were all members of the band and were therefore systematically different from the otherwise randomly comprised platoons.

Procedure. Subjects were assigned randomly by platoon to the two treatment groups and the control group. One treatment group comprised of three platoons totalling 123 cadets was instructed to keep a daily stress record. These records were to list stressful events which occurred during the day and for each event: (1) describe the event, (2) tell what the cadet did in response to the stressful event, and (3) describe the outcome of the cadet's action. Typical stressors reported by cadets were behaviors of their leaders, punishment, required attention to detail, physical exercise demands and inspections.

Responses included anger, goal setting, humor, practice, discounting and modeling. Outcomes included feelings of accomplishment, changes in performance levels and frustration. Stress diaries were collected at the end of each week and new diaries were issued. Protection of their diaries was guaranteed and feedback provided to the chain of command was promised to be anonymous. During the remainder of the summer an experimenter collected the diaries weekly and kept in touch with the cadet officers in order to encourage their support of the exercise. This treatment was intended to focus the cadet's attention on his own role in dealing with the stress of Swab Summer.

A second treatment group comprised of three platoons totalling 117 cadets received the same instructions as the first treatment group. In addition, however, this group was randomly divided into subgroups of four people. Cadets were instructed to meet weekly with the members of their subgroup and at those meetings to share one success which they had during the week. This group was called the "Meeting Group" while the first treatment group was labeled "Non-meeting Group." It was expected that the support groups would result in the treatment being more effective and, in addition, would result in the formation of significant support networks for the subjects. A control group was comprised of two platoons. All data collected for both Meeting and Non-meeting Groups was also available for the control group with the exception of the stress diaries. No instructions were given to the control group.

<u>Dependent Variables</u>.

Summer Evaluation Questionnaire. All cadets were administered a Summer Training Evaluation Questionnaire (STEQ) as part of the Coast Guard Academy regular administrative process following the completion of Swab Summer. The STEQ was not linked to this study in any way. From the 53 questions included in the STEQ, eight questions were identified as being relevant to the purpose of the study. These included:

1. Swab Summer was much more physically demanding than anticipated. (Physical Demands)
2. The most difficult aspect of Swab Summer is the psychological stress cadets must contend with. (Psychological Demands)
3. If I really knew what to expect of Swab Summer, I never would have accepted my appointment. (Regrets)
4. As promised, the training experience during Swab Summer proved to be a continuous challenge to me. (Challenge)
5. I personally would have benefitted greatly from additional free time during Swab Summer. (Need For Free Time)
6. Swab Summer was so tough I contemplated resignation almost every day. (Resignation)
7. Psychologically, Swab Summer has left me feeling strained and ill prepared for the academic year. (Strained)
8. On the average, the level of stress you experienced daily during Swab Summer was. . .none to extreme. (Stress)

Questions 1 through 7 were rated on a 5 point Likert Scale from strongly disagree to strongly agree. Question 8 was rated on a 7 point Likert Scale from none to extreme.

Academic Performance. Academic performance was indicated by the cadet's grade point average (GPA) for the five courses of the essentially identical curriculum taken by all cadets during their first semester.

Results. Results were unexpected and not only did not support the hypotheses but were exactly opposite from those results predicted. Attrition of subjects was very great. In the Non-meeting Group no subjects participated in the study for the e[illegible] ix weeks of Swab Summer. In the Meeting Group 25% of the subjects p[illegible]d for the entire six weeks. Interpreted as a significant findi[illegible] t clearly demonstrates that self-help efforts benefit from the su[illegible] peers. Other well known treatment programs such as Alcoholics Anonymous and Weight Watchers have followed this principle for some time. What is surprising is that in this military environment a commitment to share your diary with your peers is more effective than the continual "encouragement" of your seniors to participate.

The second surprise was that differences in STEQ responses were few and all significant differences showed treatment group subjects reporting higher stress than controls. Subjects in the Meeting Group, now numbering 30, perceived the summer as more physically demanding ($p < .001$), more challenging ($p < .001$) and felt that they needed more free time ($p < .01$). All other differences, though non-significant, were in the same direction with the exception of Regrets (Treatment = 1.80, Control = 1.85).

The most troubling finding, however, was the non-significant difference ($p = .06$) between the Treatment Group's GPA (2.39) and the Control Group's GPA (2.63). While this difference did not reach significance, it is doubtful that the Dean of Academics would have found that defense acceptable.

From their combined results it appeared that the combination of stress diary and support groups was effective in changing cadet perceptions of Swab Summer.

STUDY II

One reason for the findings of Study I might have been that cadets found the pressure to keep stress diaries in itself stressful. Therefore, a second stu[illegible] was planned along essentially the same lines with changes which would reduce the pressure on cadets created by the stress books. This study was conducted with 400 students in the Class of 1985 who entered the Coast Guard Academy in the summer of 1981. One platoon of 40 persons, all members of the band, was again omitted. The remaining cadets were randomly divided by platoon among the two treatment and control groups.

For the study more effort was put into the instructions concerning how to keep the records and more mention f the benefits to the cadets was made. Secondly, serious effort was made to minimize pressure to participants in the study placed on the cadets by their platoon leaders.

The results of Study III, in retrospect, might have been expected. Virtually none of the cadets in either treatment participated for the entire six weeks of Swab Summer. Again, the dropout rate was greater for cadets in the Non-meeting Group. By lowering the criterion for participation to three weeks,

29 cadets in the Meeting Group were identified as participants. Using this redefined treatment group, none of the STEQ comparisons were significant but several approached significance and all items showed treatment subjects as reporting higher stress than control subjects with the exception of one item (Challenge Treatment = 3.96, Control = 4.03). The lack of STEQ differences would suggest that there should be no GPA differences which was the case (Treatment = 2.60, Control = 2.66).

Despite the failure of these findings to reach significance, the pattern of findings supports Study I and serves as some evidence that the results of Study I were not specious. It is also useful to note that the findings decreased as the length and degree of involvement in the treatment procedure decreased.

STUDY III

In evaluating the results of Studies I and II, it became clear that given the same treatment one might have hypothesized the observed effects based on sensitization to the environment. Verbrugge (1980) in discussing her findings with respect to the keeping of health diaries notes that patients who keep symptom diaries report more symptoms than do patients who are asked to retrospectively summarize their symptoms. While Verbrugge values the increase in reported symptoms, it may also be that the keeping of health diaries acts as a placebo in reverse. The data of Study I and to some degree Study II suggest that requiring cadets to constantly attend to their stress results in a greater perceived level of stress.

Study III, therefore, followed exactly the same format with one major difference. Cadets were requested to keep logs of all the good things that happened to them instead of the stressful things. One hundred forty-eight cadets in the Class of 1986 who entered the Coast Guard Academy in July 1982 began the study. By the end of six weeks, no cadets in the Non-meeting Group and 21 cadets in the Meeting Group were participating. This pattern of participation corresponds exactly to the pattern observed in Study I.

It was predicted that participants, now the 21 cadets in the Meeting Group, would report lower stress than controls and would consequently have higher GPA's than controls. Treatment subjects, in fact, reported less psychological stress ($p <.001$), fewer thoughts about resignation ($p <.05$) and a greater readiness with less strain ($p <.01$) than did controls. All other differences showed treatment subjects reporting less stress with the exception that they experienced a slightly greater degree of physical demands (Treatment = 3.24, Control - 2.84).

GPA was taken for treatment subjects and controls at midterm and as yet no differences have appeared (Treatment = 2.52, Control = 2.54). A second indicator of performance was available for this group. Military performance derived from ratings made by peers and seniors in the cadet chain of command showed no significant differences in treatment and controls (Treatment = 646, Control = 616). Further analyses of these results pends the end of the fall semester.

Discussion

It is clear that these findings do not smite one between the eyeballs (Fiedler's noted "Intraoccular Trauma Test"). What is significant is that, given the tremendously stressful nature of Swab Summer, any treatment at all is successful. These studies were clearly not a major feature of the cadets' summer experiences, yet Study I and Study III showed significant contrasting results clearly dependent on the nature of the treatment. It does appear possible with a simple exercise to alter the way in which people perceive stressful events.

One criticism of this study can be based on the tremendous drop out rate. Self-selection surely resulted in subjects who were no longer random representatives of the cadet corps at large. A post hoc analysis using archival data which consisted of the 16PF, Edwards Personal Preference Index, and California Personality Inventory showed that cadets who persisted in the study showed higher on leadership and responsibility traits. Given this fact, it is even more significant that persistent cadets in Study I report more stress than controls and persisting cadets in Study III report less stress than controls.

While the paper may have implications for the design of future attempts to modify coping behaviors and stress perceptions of students, a more important suggestion may be relevant to everyday practice. If cadet perception can be altered with such little effort, what effect do our "rap groups" to explore problems, our critiques to solicit criticism, our constant public acknowledgement that such experiences are "horribly stressful" have on the day to day perceptions of students or service personnel? Without trying to sound like Norman Vincent Peale, leaders would do well to encourage the recognition of what works well in their organizations. Stress is truly present in our lives, but the exaggerated perception of that stress is probably a significant factor in the negative relationships that have been found between stress and performance.

References

ANDERSON, C.R. Coping Behaviors as Intervening Mechanisms in the Inverted-U Stress-Performance Relationship. Journal of Applied Psychology, 1976, 61, 30-34.

BARNES, V, POTTER, E.H. & FIEDLER, F.G. Interpersonal Stress and the Prediction of Academic Performance. Unpublished paper, 1982.

DORNBUSCH, S.M. The Military Academy as an Assimilating Institution. Social Forces, 1955, 33, 316-321.

GOPERLUND, G.N. Social Support and Stress During the First Year of Graduate School. Professional Psychology, 1980, 283-290.

VERBRUGGE, L.M. Health Diaries. Medical Care, 1980, 18, 73-95.

DEVELOPMENT OF A NEW FORM FOR TEACHER APPRAISALS OF LEADERSHIP*

Robert F. Priest, Ph.D.
United States Military Academy

ABSTRACT

Teacher ratings are not useful as predictors of academic achievement because standardized tests work better. No satisfactory standardized measure of leadership ability has been developed. Nevertheless, the military service academies are particularly concerned with development of leadership attributes and use estimates of leadership ability in selecting candidates. Though all such measures have low validities, high school teacher ratings are the most promising (Priest and Adams, 1980). This paper describes the development of a new teacher rating form which was field tested experimentally in 1980-81, and has been adopted operationally for admissions in 1981 at West Point. Preliminary data on reliability, validity, and teacher acceptance are reported.

BACKGROUND

An unstructured letter of recommendation is widely used in selection, both in industry and in education. The literature on personnel selection (Stone & Kendall, 1956; Guion, 1965) questions the predictive validity of such recommendations. A recent review of college admissions research also questions the validity of teacher recommendations in predicting performance in college (Willingham and Breland, 1982). Nevertheless, teacher evaluations of student attributes are frequently considered in selecting among applicants at competitive colleges (Aleamoni, 1972; Greenberg & O'Brien, 1976), medical school (Rainey & Luecking, 1974), law school (Pipkin & Katsh, 1976) or graduate school (Lewis, 1972). Much of the criticism focuses on the unstructured letter of recommendation because, except for Orvic (1973), almost no work has been done on the predictive validity of structured ratings made in the context of recommendations. Present research compares the reliability and validity of two systems of teacher ratings.

Standardized tests are superior to teacher ratings as measures of cognitive abilities needed in college (Cleary et al., 1975; Stanley, 1976). However, it is often asserted that teacher ratings can be useful in predicting noncognitive criteria such as motivation, creative accomplishment, or leadership. No standardized test has proved to be satisfactory in predicting leadership. Given the importance of selecting and training leaders in the Military, the Service Academies have taken care to develop systematic measures for assessing the leadership potential of high school students. For many years, the U.S.

---

*Any conclusions in this report are not to be construed as official U.S. Military Academy or Department of the Army positions unless so designated by other authorized documents. The author wishes to thank Richard Butler and Carlton Bacon for suggestions on earlier phases of this project.

Military Academy has used three elements to quantify the "leadership potential" of candidates: an athletic accomplishment score, based on a biographical record of student participation & achievement; an extracurricular accomplishment score; and a Faculty Appraisal Score (FAS), based on ratings from four faculty members with diverse perspectives--guidance counselor, coach, English and math teachers. Research studies over a period of over 17 years have shown low validity coefficients for all such measures; but FAS has consistently been the most promising predictor (Priest, 1980). Even though validities are low, there is justification for using such measures in selection (Schmidt et al., 1979). Recent work showed that the FAS was not biased against any one race or gender (Priest & Adams, 1980). Because of the importance of leadership as a criterion and the promise of earlier FAS measures, the Academy began a project to improve the FAS. This report compares the new rating system to the one in use in 1979-80.

METHOD

In 1979-80, the Military Academy developed a new system for appraising the military performance of cadets. Whereas prior appraisal systems focused to some extent on future potential, the new system was based on behaviorally anchored rating scales of six dimensions of current military performance: task structuring & management, interpersonal relations, compliance with organizational expectations, intellectual application & growth, personal & professional ethical behavior, and performance oriented development.

In 1980, USMA began to develop a new Faculty Appraisal Form which would meet several criteria: (1) it should reflect the six dimensions of military performance; (2) it should be acceptable to teachers in the field; and (3) it should have satisfactory inter-rater reliability and predictive validity.

A committee of officers and staff generated over 160 items to fit these six dimensions. They were given a list of ten criteria for evaluating items, and on the basis of their preliminary ratings, the 24 items with the best mean scores were selected for further development. Two alternative rating formats were developed and evaluated by a sample of 27 high school guidance counselors. The most acceptable rating format was designed to reduce rating inflation. Comments by counselors were also used to edit and revise a final set of 15 items (see Table 1).

The new form was field-tested with the Class of 1985. The original research plan specified giving two old 10-item forms and two new 15-item forms to each candidate, so that every pair of raters (from the set of Guidance Counselors, and English, Math and PE teachers) would be equally represented. Unfortunately, the new forms arrived two weeks late from the printer, and the first group of candidates received only old forms. The admitted class file contains ratings for 796 cadets who were rated only on the old form, 126 who were rated only on the new form, and 760 who were rated on both. Of those rated on both, 465 received an equal number of old and new rating forms.

In late 1980, an in-process review of the new form was conducted. The new form was well-accepted in the field by teachers who used it. Military Academy admissions officers liked the greater specificity of comments on the new form. Institutional Research examined 101 candidate folders where there was at least

TABLE 1

THIS CANDIDATE HAS DEMONSTRATED AN ABILITY TO:

1.	Make friends easily	B
2.	Show interest and concern for the welfare of others	B
3.	Influence other students to work together	B
4.	Communicate effectively in face to face discussion	B
5.	Communicate effectively in written work	B
6.	Set an example of good conduct for other students	C
7.	Exert maximum effort, showing a strong desire to achieve in every field	C
8.	Show self-control and perform well, even under pressure	C
9.	Adjust to a demanding schedule of activities without neglecting school work	A
10.	Set high standards for own performance in a number of areas of school work	A
11.	Seek academic challenge beyond that required by normal coursework	D
12.	Accept criticism and make improvements from it	E
13.	Accept full responsibility for personal shortcomings	E
14.	Teach practical skills to others	F
15.	Correct others who make mistakes in firm but supportive manner	F

Notes: B = Interpersonal relations, items 15.
C = Compliance with organizational expectations, items 6-8.
A = Task structure & management, items 9-10.
D = Intellectual application, item 11.
E = Personal & Professional Ethical Behavior, items 12-13.
F = Performance oriented development, items 14-15.

one old and one new form, and discovered the new form was having the intended effect of reducing rater inflation: on the old form 56% of the ratings were in the top block ("superior"), in contrast to 29% on the new ("top 1%"). Although the planned study of reliability and validity of the new form was not scheduled to be completed until one year after the Class of 1985 had entered (i.e., 1982), the Admissions Office decided on the basis of the in-process review (and for administrative simplicity) to adapt the new form for operational use for the Class of 1986.

RESULTS

Table 2 shows the correlations between total scores by different teachers rating the same candidate. Both old and new forms were equally low in inter-rater reliability. The new form was more reliable for certain rater pairs, but the old form was more reliable for other rater pairs. Another analysis (Priest, 1982) shows that more of the <u>items</u> on the new form had inter-rater r's greater than .17 than on the old form.

Table 3 shows the validity of the old and new forms in predicting first semester Military Development Ratings. The new form is slightly more valid for four of the six rater pairs. Although it appears that different rater pairs have different validities, in fact, there is no significant difference among the six validities for the new form [$\chi^2(5) = 5.3$, $p > .05$]. For cases with an

TABLE 2

INTER-RATER RELIABILITIES

Rater-Pair	New Form N	New Form r	Old Form N	Old Form r
English-Math	79	.20	423	.33*
English-Guidance	142	.27*	496	.14*
English-Coach	139	.22*	428	.18*
Math-Guidance	84	.30*	396	.19*
Math-Coach	112	.10	374	.22*
Guidance-Coach	65	.10	367	.26*

TABLE 3

VALIDITY IN PREDICTING FIRST SEMESTER MILITARY DEVELOPMENT GRADE

Rater-Pair	New Form N	New Form r	Old Form N	Old Form r
English-Math	59	.24*	384	.15*
English-Guidance	119	.15*	446	.11*
English-Coach	114	.27*	380	.14*
Math-Guidance	68	.27*	359	.13*
Math-Coach	90	.14	337	.18*
Guidance-Coach	54	-.07	324	.03

*$p < .05$

NOTE: Using all cases with at least two raters on each form and no more than 7 (new form) or 5 (old form) items marked "unable to judge."

equal number of old and new ratings, the new form has a validity of .123, in contrast to .115 for the old form, disregarding the source of the ratings.

There is some evidence that the new military development rating system is more oriented to academic achievement than former leadership rating systems were. For example, the academic predictor used by USMA (a combination of High School Rank and College Board Test Scores) correlates .20 with Military Development Ratings ($N = 1,341$, $p < .001$). That is, the USMA academic predictor is a better predictor of "leadership" than faculty ratings on old or new forms. In former years, Physical Aptitude Test Score was also a modest predictor of leadership ratings, but in the current sample, its validity is only .08.

DISCUSSION AND CONCLUSIONS

The new form is quite acceptable to users, but its reliability and validity is only slightly better than the old form. There are several limitations and constraints in evaluating these results. Both practical experience and psychometric theory show that ratings of any type tend to be more reliable and valid when the ratings of several different observers are combined or averaged. In the present application the reliability and validity of the ratings on the new form are based on, at most, two raters. Thus, their reliabilty and validity are likely to be misleadingly low. It is expected that the validity coefficients reported here will be larger when complete data from four or five raters are available for the Class of 1986.

All the admitted cadets, on whom this analysis is based, were carefully screened for leadership potential. Thus, even though the new form was not explicitly quantified for use in admissions, we assume that candidates were selected partly on the basis of how they were rated on the new form. As a consequence of this restriction in range, the validity coefficients are lower than they should be, and most importantly, tend to understate the true importance of the FAS in selection. If USMA were to deliberately admit a few candidates with the lowest possible FAS (for example--an obnoxious, but brilliant scholar, or a prize athlete who was thoroughly disliked by his teachers) to achieve class balance goals, and if the form is as good as we think it is, it would have a marvelous effect on raising the validity coefficients.

Little research has been done to quantify the impact of unstructured teacher comments on the admission process at USMA. Given the amount of time spent by admissions officers in reading such comments, such research may prove worthwhile.

The new form was explicitly designed to measure six dimensions of military performance. Some dimensions may prove to be more important or more predictable than others. Thus, further research is needed to ascertain the convergent and discriminant validity of the six dimensions.

The criterion variable, end-of-first-semester leadership grades, was chosen in part because it provided an opportunity for relatively quick feedback on the validity of the new form. Nevertheless, we plan to continue to study the leadership of these cadets in later years, in order to ascertain the generality of the initial validities reported here.

REFERENCES

Aleamoni, L. The vagaries of selection--1. Selection, placement and proficiency in American colleges and universities. The National ACAC Journal, 1972, 17, 7-12.

Cleary, T A., Humphries, L. Kendrick, S., & Wesman, A. Education uses of tests with disadvantaged students. American Psychologist, 1975, 30, 15-40.

Greenberg, A., & O'Brien, A. College admissions 1975. Paper presented at the annual meeting of the American Personnel and Guidance Association, Chicago, 1976. Abstracted in ERIC, ED128671.

Guion, R. Personnel Testing. New York: McGraw-Hill, 1965.

Lewis, L. On the Genesis of Gray-Flanneled Puritans. AAUP Bulletin, 1972. 58, 21-29.

Orvick, J. The predictive validity of premilitary performance ratings by high school personnel, Journal of Applied Psychology, 1973, 57, 88-89.

Pipkin, R., & Katsh, E. Undergraduate legal studies and law school gatekeepers. Journal of Legal Education, 1976, 28, 103-111.

Priest, R. F. Reliability, validity, and bias in measures of high school leadership (#80-006). West Point, NY: USMA, Office of Institutional Research, May 1980.

Priest, R. F. Initial validity of a new school official's evaluation form (#USMA-ODIR-82-006). West Point, NY: Office of Institutional Research, June 1982.

Priest, R. F., & Adams, J. Validity, not bias, in teacher ratings of leadership. Paper presented at the American Psychological Association, in Montreal, Canada, 5 September 1980.

Rainey, M., & Luecking, D. Admissions committee gatekeeping: the case of the female applicant to medical school. Paper read at the Midwest Sociological Society. Omaha, April 1974. Abstracted in College Student Personnel Abstracts, Spring 1975, 10, 284-285.

Schmidt, F., Hunter, J. E., McKenzie, R., & Muldrow, T. Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 1979, 64, 609-627.

Stanley, J. Test better finder of great math talent than teachers are. American Psychologist, 1976, 31, 313-314.

Stone, C H., & Kendall, W. F. Effective personnel selection procedures. Englewood Cliffs, NY: Prentice-Hall, 1956.

Willingham, W. & Breland, H. Personal qualities and college admissions. New York: College Entrance Examination Board, 1982.

AD P000899

Use of Delphi-Based Interdisciplinary Goals for Evaluating Course Outcomes*

Robert F. Priest, Ph. D.
U.S. MILITARY ACADEMY

A. INTRODUCTION

Over the past decade there has been a renewed interest in the teaching of ethics. Starting in 1977, the Hastings Center for Society, Ethics, and the Life Sciences began a study of the teaching of ethics in American higher education (Stromberg, Wakin, and Callahan, 1982). At about the same time, Department of the Army was studying ways to make the overall structure and environment of the US Military Academy (USMA) more conducive to the moral growth and development of cadets. In 1977 a Department of the Army study recommended that USMA "establish a comprehensive and progressive program in ethics and professionalism to prepare cadets for the ethical...problems that confront officers." The study further recommended that USMA establish a committee to insure that the program was integrated into the USMA curriculum (Dickson, et al., 1977). Starting in 1978-79, cadets were required to take eight courses which related to ethics and professionalism: Military Heritage/Standards of Professional Behavior, Behavioral Science, Philosophy, Law, History of the Military Art (that is, military history), Military Psychology and Leadership, and an interdisciplinary course in American Institutions. Three of these courses were new courses (Military Heritage, Philosophy, and American Institutions). The remaining courses were established courses which provided important components of the cadet's professional and ethical instruction.

In order to integrate USMA's academic program with the honor code, religious activities, and other aspects of the Academy which promote moral or professional development, a permanent interdisciplinary Ethics and Professionalism Committee was established in 1979. The committee reviews and evaluates programs and occasionally publishes papers to stimulate faculty dialogue on ethical and professional matters. One of the committee's tasks - evaluating the teaching of ethics-is particularly difficult and controversial, even for a single course, let alone a broadly based interdisciplinary group of courses. Nevertheless, to evaluate progress in meeting goals requires some attempt at measuring student reactions to the program (Stromberg et al., 1982, p. 51 - 55).

Few colleges attempt to structure interdisciplinary programs such as USMA has done with its ethics and professionalism "curriculum", and few have attempted to evaluate the cummulative result. Dressel (1976) notes that most undergraduate curricula are lacking in sequence and direction. When output objectives common to a set of courses are not specified, the result is a "pre-occupation with specific knowledge" (p.303). The purpose of the current study was to provide input into the development of an end-of-course critique sheet for the eight

*Note - Any conclusions in this paper are not to be construed as official U.S. Military Academy or Department of the Army positons unless so designated by other authorized documents. The author wishes to thank Richard Butler and Carlton Bacon for suggestions on earlier phases of this project.

Ethics and Professionalism-related courses, one which would provide reliable, objective and useful information. If such a common measuring standard can be developed, it would increase the likelihood that the Academy would be able to measure whether or not the courses were helping the cadets grow over the four years in cognitive skills and attitudes relevant to ethics and professionalism. It would also promote teamwork among courses on those issues where it is appropriate, while at the same time respecting the autonomy of each course to pursue its own unique goals.

B. METHOD

The Delphi technique is a method for developing and improving group consensus (Anderson, et al., 1976). It is an interactive process, involving a group of experts in formulating goals for policies and coming to agreement about them through successive stages using a questionnaire. Although the Delphi technique originated in the context of technological forecasting (Linstone and Turoff, 1975), it has been used to formulate national policy on drug abuse among U.S. experts (Jellison, 1975), to define competency and educational objectives in podiatric medicine (Lanham, 1979), and curriculum planning for undergraduate chemistry (Melton, 1977).

The general procedure in the present study applied the Delphi technique as follows: it asked instructors to respond to a list of proposed common goals, and to revise them or add new goals as desired. Next, the researcher summarized the responses of instructors, provided written feedback on the results of the survey, and administered a second Delphi survey to the original instructor group. Finally, the results of the second Delphi survey were summarized and sent to instructors to provide feedback to them, as promised.

A list of course objectives for each of the eight ethics and professionalism courses was supplied by the eight course directions. Many objectives, if not most, were stated in terms of particular facts, knowledge, or cognitive skills which are unique to each particular course. Since the objective of this project was to discover the general principles and skills common to all eight courses, it was evident that many of the statements would have to be restated or reformulated in more general terms so as to be more generally applicable.

Both Callahan (1980) and Caplan (1980) have discussed the problems of formulating goals in ethics teaching and evaluating such goals. Based on Caplan's work and on specific course objectives, a list of 24 possible common goals was formulated.

Each of the 24 statements represent an outcome that instructors strive for, or wish to avoid, in their course. Based on the work of Peterson (1970) and others, we decided to inquire both into the perceived existing goals (to what extent does the course you teach, as you taught it, emphasize attainment of this criterion") and the ideal goals ("to what extent should it").

A six category response code was used for all items: "Maximum possible emphasis" (50 points), "extremely strong emphasis" (40), "strong emphasis' (30), "moderate emphasis" (20), "little emphasis" (10), and "no emphasis" (0). Prior work showed that instructors could employ these categories in a discriminating manner to describe cadet performance.

The original phase 1 questionnaire was administered by the Office of the Dean to all instructors who had taught one of the eight courses one complete semester. The Delphi method usually identifies a group of experts, and in this application, one semester of teaching experience was required to qualify as an expert in course goals. Based on analysis of phase 1 data, six statements were dropped and 13 instructor-generated statements were added to the phase 2 questionnaire. This report is based mainly on an analysis of phase 2 results.

Table 1

Number of Instructors in Each Phase

Course		Phase 1	Phase 2
MS 101	Military Science	9	20
PL100	Intro. Behavioral Science	6	7
PY201	Ethics	8	9
LAW 300	Law	6	9
HI300	History of the Military Art	9	4
HI383/HI302	History of the Military Art*	2	2
PL300	Mil Psychology & Leadership	4	5
AI479	American Institutions	2	1
		46	57

* Two Semester Course

To evaluate whether or not a given goal was seen as equally important in all courses, a oneway analysis of variance was computed for each item. If the mean emphasis on a goal in the eight courses was sufficiently different, a statistically significant test statistic resulted. In this analysis, we concluded that items with no statistically significant difference among courses would be considered "common" items. In this analysis, "common" items are not necessarily the items with the highest overall mean emphasis ratings.

C. RESULTS AND DISCUSSION

The phase 2 questionnaire asked: "Are there some statements which you would consider abstract platitudes with little relevance for day to day teaching? If so, list the item numbers below." When four or more instructors nominated the same item, it was considered a platitude. Three items out of the 31 were considered platitudes: these items also have the lowest mean importance rating. Thus, "abstract platitudes with little relevance for day to day teaching" were given low emphasis by instructors in actual practice as well as in ideal emphasis.

Table 2 lists the items which qualify as common goal items because instructors in the eight courses do not differ significantly in their perception of the actual emphasis or the ideal emphasis for that item. The last two items are also regarded as platitudes. Thus, there are 12 items in the "strong" or "extremely strong" emphasis category which are not platitudes, and can serve as the basis for a future end-of-course critique.

Table 2

List of "Common Goals"

Interpretation Of Mean Score	
Extremely strong emphasis	Participate actively in classroom discussions, where appropriate.
Extremely strong	Believe the course content is relevant to broad professional issues outside the classroom.
Extremely strong	Refer to course readings and lecture materials in class discussions.
Extremely strong	Accept personal responsibility for moral ethical behavior.
Extremely strong	Demonstrate an ability to analyze behavioral, historical, legal, military or philosophical issues increasingly well over the semester.
Extremely strong	Are able to conduct a reasonably coherent discussion about a professional or ethical issue of importance.
Extremely strong	Are able to identify the potential ethical and professional issues in a given case study, hypothetical example, historical account, or legal procedure.
Strong	Go beyond the lectures and readings in applying ethical and professional issues.
Strong	Raise issues concerning moral or professional problems in appropriate circumstances outside the classroom.
Strong	Continue their study and consideration of ethical issues after the course ends.
Strong	Use in spontaneous informal discussions with classmates concepts and theories of ethics and professionalism developed in the classroom.
Strong	Develop their own individual moral philosophy and are able to defend it rationally.
Moderate	Are able to express why it is that loyalty to the Military Academy is an essential trait.
Moderate	Come to a clearer understanding of why Academy rules and regulations are as they are and why their superiors decide as they do.

One of the goals of the Delphi method is to build and improve group consensus. It would be expected that the standard deviation of ratings for specific goals would be smaller on phase 2 than for phase 1. This hypothesis was tested for the 18 items which were in both the phase 1 and the phase 2 questionnaire. For the actual emphasis ratings, 10 of the 18 standard deviations decreased; for the "should" ratings, seven of the ratings decreased. Overall, these results do not indicate a trend toward greater consensus among the instructors as a group for all 18 items. Furthermore, there was no trend toward increased consensus within courses. Although the Delphi technique did not lead to increased consensus in this application, it did identify the items of greatest common interdisciplinary agreement (Priest, 1982, Note 1).

D. CONCLUSIONS

The first 12 statements in Table 2 represent the general goals which are common to the eight ethics and professionalism courses, and are also given sufficient emphasis to warrant inclusion.

An evaluation of a particular course should be broadly based on both a course-specific component and an interdisciplinary component. Thus, the present research study provides only one component, the interdisciplinary one, for a complete course evaluation. Almost all previous work in developing measures of student evaluators of courses has focused either on goal-free assessment (i.e., not specific to course objectives) or only on the goals of a particular course. The present study raises the interesting possibility of studying relationships between attainment of course-specific objectives and the attainment of interdisciplinary objectives. We are currently conducting follow-up research to measure cadet perceptions of both course emphases (actual and ideal) and of their own attainments in each course. If the interdisciplinary ethics and professionalism program is aiding cadet development, we would expect to find improved self-perceptions of cadet attainments on these interdisciplinary objectives over the four year program.

References

Anderson, S., Ball, S., Murphy and Associates, Encyclopedia of Educational Evaluation, San Francisco, CA: Jossey Bass, 1976.

Callahan, D. "Goals in the Teaching of Ethics" in D. Callahan and S. Bok (Eds.) Ethics Teaching in Higher Education, Plenum Press, New York, 1980.

Caplan, Arthur L. "Evaluation and the Teaching of Ethics" in D. Callahan and S. Bok (Ed.) Ethics Teaching in Higher Education (The Hastings Center Series in Ethics), Plenum Press, New York, 1980.

Dickinson, H., Mackmull, J.V. and Merritt, J.N. Final Report of the West Point Study Group Department of the Army, Office of the Deputy Chief of Staff for Personnell, Washington, D.C. 20310, 27 July 1977.

Dressell, P. Handbook of Academic Evaluation, San Francisco, CA: Jossey Bass, 1976.

Goodpaster, A. Basic Concepts For The United States Miliatary Academy, West Point, NY: February, 1979

Jellison, I. "The National Drug Abuse Policy Delphi: Progress Report and Findings to Date", In Linstone and Turoff, M. (Eds.) The Delphi Method, Techniques and Applications, Addison Wesley, 1975.

Lanham, R. Clinical competency in Podiatric Medicine, Journal of Podiatric Medical Evaluation, 1979, 10, p38-40.

Linstone, H. and Turoff, M. (Eds.) The Delphi Method, Techniques and Applications, Addison Wesley, 1975.

Melton, IA A study of undergraduate education in chemistry by the delphi technique Journal of Chemical Education, 1977, 54, 640-1.

Peterson, R. The Crisis of Purpose, Definition and Uses of Institutional Goals, Report No. 5, prepared for the ERIC Clearing house on Higher Education, October 1970.

Stromberg, P., Wakin, M. and Callahan, D. The Teaching of Ethics In The Military. The Hastings Center, 360 Broadway, Hastings-on-Hudson, NY 10706: 1982.

Reference Note

1. Priest, R. Use of Delphi-Based Interdisciplinary Goals for Evaluating Course Outcomes. West Point, New York; Office of Institutional Research Report 82-003, April, 1982.

AD P000900

JOB DESCRIPTIONS DEVELOPED FROM GROUPS X MODULE (GXM) ANALYSES

Stephen S. Roop

Texas Transportation Institute
Texas A&M University
and
Richard W. Dickinson

Occupational Research Division
Industrial Engineering Department
Texas A&M University
College Station, Texas 77843

INTRODUCTION

Hierarchical clustering as defined by Ward (1963) is at the heart of the task inventory approach to job analysis. It is the results of the grouping process that many occupational anayses use to evaluate characteristics of interest in a given population. Clustering, stated in the most basic terms, is an iterative process that combines, in stages, the most similar of a given array of objects. As most commonly used in occupational analysis, clustering forms a series of groups based on the similarities between incumbent relative time spent ratings. The similarities in incumbent relative time spent ratings are determined, when using absolute overlap, from two features of the incumbent response. The first feature is what can be called the "pattern" of the response. That is, those specific task statements responded to be by a given incumbent that are representative of that incumbent's job. The second feature is the "magnitude" of the time spent ratings. Absolute overlap is calculated by summing the minimum common value between corresponding responses for two incumbents. Corresponding responses referring again to the pattern of response.

Traditional job descriptions are comprised of the average time spent ratings for a group of similar incumbents. Every task responded to by a group member will appear in the group's job description. As a result, many task statements appear in the job description that are not characteristic of the group's similarity. In fact, the frequency of these outlying task statements increases as group homogeneity decreases.

Typically the task statements within group job descriptions are sorted in descending time spent order to allow the job analyst to focus on those tasks that are representative of group commonalities in the distribution of work time. For many purposes this is a productive approach.

An alternative method presenting the data is the listing of task

statements in the job description by a previously defined "duty field".

Duty fields are typically comprised of tasks that have some feature or characteristic in common. These features are defined in an apriori fashion and thus do not address an attribute of tasks that has some importance to occupational analysis. That attribute is the "occupational relatedness" of tasks. Occupationally related tasks are performed to a significant extent in concert with one another and thus define the interrelationship of tasks in addition to defining job content.

This paper details one possible approach to defining and describing jobs with occupationally related task modules. The advanced processing required for the analysis was made possible by the enhanced processing capabilities of the CODAP80 System developed at Texas A&M under sponsorship by the U.S. Navy. The occupational data used in the analysis was obtained from the responses of 283 Navy Minemen to 232 task statements.

THE GROUPS X MODULE (GXM) APPROACH

The method selected to determine the degree of occupational relatedness in this analysis was a binary overlap algorhythm discussed by Phalen, (1981):

$$\frac{@}{A + B - @} \qquad (1)$$

where @ = number of incumbents performing task A & B

A = number of incumbents performing task A

B = number of incumbents performing task B

Clustering tasks using binary overlap identifies the pattern of incumbent response. If incumbents perform task 1, this measure indicates the extent that they also perform task 2 or task 3, etc. Task clustering can be thought of as a process that provides a dual solution to the incumbent clustering. It should be stressed that the result of the binary clustering highlights the pattern of response only; magnitude of response, which is captured directly in the incumbent clustering, is evaluated at a subsequent stage in this analysis.

The results of the binary clustering on tasks is used to define task modules. Task modules are analogues to incumbent groups in the clustering process. The selection of modules, however, differs significantly from that of groups in that groups, once selected, remain constant. The groups represent stable bodies in the occupational setting and efforts are directed toward describing the characteristics of the group.

Task modules, on the other hand, are under no such restriction. During the iteritive process of clustering, modules form based on the degree of incumbents' inconcert performance. As the collapsing process progresses,

modules with the most similar performance-nonperformance characteristics are brought together. The result is a distinctive array of task module "families" that represent general patterns of task relatedness in the occupational setting.

The clustering process procedes from the formation of modules with a high degree of relatedness to the combination of modules with a lower degree of relatedness. By evaluating a group's time spent values across an entire family of modules a precise description of the group may be obtained with only a few modules and a minimum in lost information.

A module family is comprised of basic elemental units termed "base" modules. The combining of base modules in the form defined by the collapsing process allows the analyst to account for every task statement in the inventory in addition to expressing the group's job description in the most general terms possible. This strategy enables the job description to retain the characteristic of occupational relatedness.

The selection of a given module for a group under investigation is guided by the degree of time spent "over-representation" calculated for the group and module in question. The index used for this determination is the "core ratio".

CORE RATIO

The results of the incumbent clustering on relative time spent using absolute overlap is combined with the results of the task clustering on incumbent performance-nonperformance to yield a Groups X Module (GXM) matrix. The core ratio is used to unitize the magnitude of time spent in each module of tasks in order to effectively select those modules that are representative of a group's work time distribution.

The core ratio is formulated as follows:

$$CR_{ij} = (T_{ij}/100)/(n_j/N) \quad (2)$$

where CR_{ij} = core ratio for group i for module j

T_{ij} = time spent value for group i in module j

n_j = number of tasks in module j

N = total number of tasks in inventory

A critical value for the core ratio is derived based on two assumptions. The first is that incumbent responses to the task inventory represent, in sum, 100 percent of an incumbent's work time. The second assumption is that, prior to the analysis of incumbent responses, the probability of response to a given task statement is equal to the probability of response to any other task statement. This second assumption provides a convenient procedure for gauging the magnitude of incumbent response across all occupational subgroups. Thus, the critical value of the core ratio is defined as:

$$\text{CRITICAL CR} = 1.0$$

Core ratio values in excess of 1.0 may be taken as evidence of time spent over-representation and the module in question is thus selected for use in the job description.

QUALITY OF GXM JOB DESCRIPTIONS

Extensive comparisons performed between the task statements listed in traditional job descriptions vs. the task statements listed in the GXM job description confirm the notions underlying the GXM selection criteria. The results show that a significant portion of the group job description is preserved while the extraneous or uncharacteristic task statements accounting for a small percentage of time are removed. A sample of 10 groups selected from the Navy Mineman study and evaluated with the GXM approach revealed that, on the average, 89.7 percent of the group's time was described in modular form with 64.1 percent of the task statements from the traditional job description. This, of course, is not a significant finding in itself. Sorting task statements by time spent would enable a similar reduction in job description length. The significant contribution of the GXM approach is the assemblage of occupationally related task statements into modules based on quantitative criteria. Arrangement of tasks in this fashion allows the analyst to easily determine, describe and evaluate the characteristics of the work settings under investigation.

The following figures present a more detailed evaluation of a single group's job description from data derived from the Navy's Mineman study. Group 179 (G179) has 10 members and a traditional job description of 124 tasks. Of these 124 tasks the GXM job description retained 71 tasks in modular form and rejected, as not characteristic of G179, 53 other tasks. Figure 1 shows the frequency or accepted vs. rejected tasks at the various percent performing levels within G179.

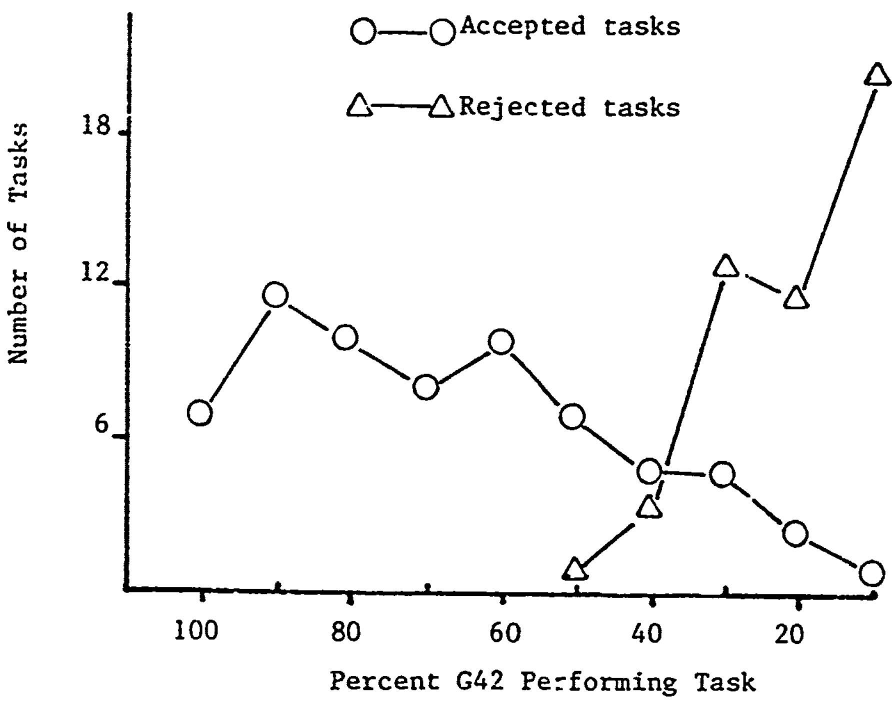


Figure 1 Number of tasks accepted vs number of tasks rejected with GXM Jobdec at different percent performing levels.

As can be seen in Figure 1, the accepted tasks are those performed by a majority of G179 members. Those tasks uncharacteristic of G179 are rejected.

Figure 2 shows the per task time spent values for accepted vs. rejected tasks across the percent performing values for the same group. The GXM approach effectively screens out those tasks that do not account for a significant portion of the group's time.

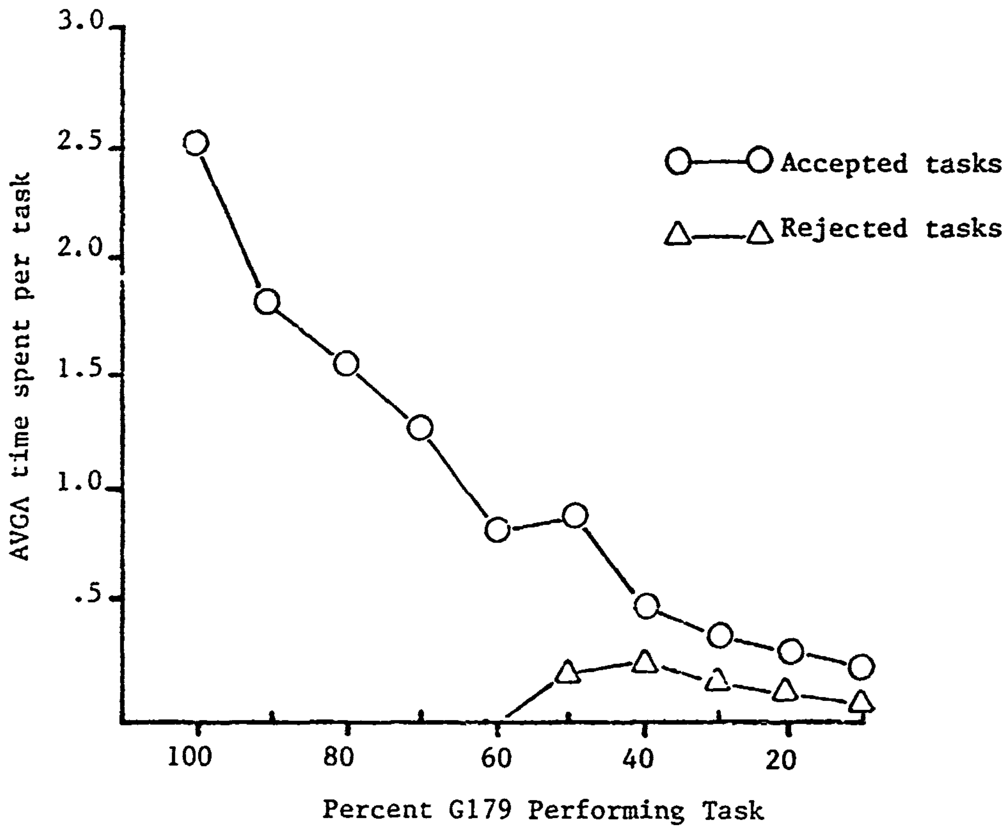


Figure 2 AVGA time spent per task at different percent performing levels for accepted vs rejected tasks.

The overall amount of time spent accounted for by the accepted tasks for G179 was 87.93 percent. This general pattern of acceptance and rejection of tasks has been found to hold for all groups tested suggesting that the core ratio measure operates in the expected manner.

CORE RATIO INDEX AND WITHIN GROUP HOMOGENEITY

Formula 2 presented the calculation used to determine the degree of

time spent over-representation of a given GXM cell. The same general formulation may be applied to the sum of the module selections to measure the coherence of the final job description. The core ratio measure may be calculated for both the accepted (CR_a) and the rejected (CR_r) tasks for a given group. This measure may be called for the purposes of this paper, the composit core ratio.

$$CR_a = (\sum_j T_{ij}/100)/(\Sigma n_j/N) \qquad (3)$$

$$CR_r = (\sum_t T_{ir}/100)/(n_{ir}/N) \qquad (4)$$

where: CR_a = core ratio for group i for all selected modules

CR_r = core ratio for group i for all rejected (r) tasks

T_{ir} = sum of the time spent value for group i for rejected tasks in traditional job description (t)

n_{ir} = number of rejected tasks from traditional job description for group i.

The composit core ratio for accepted tasks is a unitized value representing the degree of time spent over-representation for a given group on its selected task modules. The composit core ratio for rejected tasks may be interpreted as the degree of time spent under-representation on tasks that are uncharacteristic of the group as a whole. Both measures rely on the allocation of 100 percent of a groups work time across a set of defined task statements; the traditional job description. The fact that job descriptions vary in length means that a varying percentage of time spent is allocated to tasks, from group to group, as a function of job description length. The comparison of proportional time spent to the total number of task statements in the inventory (N) thus tends to bias the measure downward for groups with long job descriptions and upward for groups with short descriptions.

An effective control for the systematic bias introduced by varying job description length is obtained by evaluating CR_a in relation to CR_r for each group. The ratio of these two measures, the core ratio index, has been found to yeild a good measure of the quality of the GXM job description. The core ratio index is defined as:

$$\text{CORE RATIO INDEX} = CR_a/CR_r \qquad (5)$$

An evaluation of the core ratio index across groups selected from the Navy's Mineman study shows that the CR Index correlates highly (r=.95) with the "Within" group homogeneity figure (n=10 cases). This result suggests that as group homogeneity increases the quality of the GXM job description improves. While this is not an unexpected finding, it does lend further support to the contention that the GXM approach is capturing and describing the fundamental patterns of task performance in the occupational setting under investigation. The predictive capability of Within group homogeneity with respect to the core ratio index holds, in the present study, for groups of $N \leq 16$ incumbents. For groups larger than this, the Within group homogeneity

figure fails to correspond directly to the increases in the core ratio index. This limiting feature of the Within value was pointed out in a recent paper by Phalen and Weissmuller, (1981) in which they discuss the need, and present a method for improved job type identification. The method presented to overcome some of the difficulties encountered in previous measures of group homogeneity is the core-task homogeneity index. This measure concentrates on those tasks that are most representative of a specific group of workers and the amount of time the group devotes to those tasks.

As with the GXM job description, the core-task homogeneity index ignors the many tasks that are specific to individuals and are thus not truly characteristic of the group as a whole. It is a measure that should provide the job analyst with an improved method of group selection. The optimization of group selection in conjunction with the GXM job description offers a powerful tool to the occupational analyst. It should allow the identification and description of the various groups and subgroups within a population with increased accuracy by capturing the underlying property of occupational relatedness.

The CODAP80 system, which makes the processing requirements for the GXM approach manageable, also has the capability to produce the resulting job description in modular form. It allows the analyst to specify how the task statements are arrayed within modules and can calculate descriptive statistics either across all selected modules or within selected modules. Since task modules often combine task statements from a number of pre-defined duty fields, these labels, in addition to others may accompany the task statements.

This paper does not attempt to explain in detail every step in the GXM approach. It does attempt to lay-out the general strategy and premises as developed to date. The potential applications of the GXM approach are many and varied. Training, succession planning, employee compensation, to name a few, may all benefit from the definition of task performance patterns.

REFERENCES

Phalen, William J., Weissmuller, Johnny, J. CODAP: some new techniques to improve job type identification and definition. Paper presented at the Military Testing Association annual meeting, Washington D.C., 1981.

Ward, Joe H., Jr. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 1963, 58, 236-244.

AD P000901

Using Rating Scales to Determine Aptitude Requirements of Army Systems

Paul G. Rossmeissl and John A. Dohme
U S Army Research Institute

The United States Army is facing the possibility of a serious manpower shortage in the not to distant future. Three factors, operating concurrently, are contributing to this shortage. First, census data indicate that the quantity of individuals available for military service (18-25 year olds) will decline throughout this century and, if the birth rate remains unchanged, for the foreseeable future. Also, standardized aptitude and achievement test scores have shown a consistent decline over the past 15 years (Waters, Eitelberg, & Laurence, 1981). Taken together, these two factors indicate increased future competition among the armed forces and the civilian sector for qualified personnel, with the competition expected to be most severe for the more highly skilled individuals.

The third factor is the increasing technological sophistication of the Army's new systems. It is widely accepted that increased sophistication is increasing operator and maintainer job complexity and in turn increasing skill requirements and quantitative demand for personnel (Kerwin, Blanchard, Atzinger, & Topper, 1980), although quantitative evidence of this suspected trend is lacking (GAO, 1981). The Army, therefore, faces the possibility of increasing quantitative and qualitative personnel demands while the capability of the population to fill that demand is decreasing.

This specter of manpower shortage makes it all the more important that the Army investigate and develop techniques that will help make optimal use of the personnel that are available.

This paper reports the results of a study to assess the feasability of using rating scales to estimate the aptitudes or abilities required to operate and maintain Army systems. If accurate aptitude estimates can be obtained in this manner, the methodology could prove to be useful in two manners. First, the scales could be used to estimate the aptitude requirements of Army systems still in the design process (Rossmeissl, Kosfyia, and Baker, 1981). Second, aptitude requirement information from systems about to be, or already, fielded could be used to develop selection and classification instruments to assist in the assignment of personnel to jobs.

The current research investigated three aspects of the utility of obtaining estimates of army aptitude requirements using rating scales. If rating scales are to be useful in the context they should show three properties: they should have high inter-rater reliability, they should reliably discriminate among the aptitudes being investigated, and they should discriminate among different Army jobs.

Method

Rating Scale Development. Army aviation was selected as a test bed for investigating the use of rating scales, so a set of scales were developed that would be directly relevant to four Army helicopter missions: Aeroscout, Attack, Cargo, and Utility.

Using task analysis procedures thirty aptitudes or abilities were identified as being possible requirements for the helicopter missions. One rating scale was then developed for each of those aptitudes. The final rating scales were very similiar to those used by Fleishman (1972, 1975) in that each scale contained the Fleishman definition of the aptitude and a seven point linear rating scale. The current scales did differ from those typically used by Fleishman, however, in that the scale anchor points were directly relevant to Army aviation.

To develop these aviation-specific anchors for the 30 abilities, an ARI psychologist and an ARI Master Aviator developed as many Army aviation task statements as possible for each aptitude. The objective was to create anchor candidates that would cover the range of each aptitude from the least to the greatest amount required in performing all four Army aviation missions. In other words, to develop mission general statements that would be common to all four missions. For each ability, 1520 candidates anchor statements were generated using the Aircrew Training Manuals (ATMs) and helicopter Operator's Manuals (-10's) as guides.

Once the anchor candidates were generated, two Standardization Instructor Pilots (SIPs) were brought in to represent each mission and a roundtable discussion was held to eliminate those candidate statements that did not apply to all four missions. Certain mission oriented candidates were also eliminated because they were not part of the training regimen for a given mission. In addition, the eight SIPs edited the working of the candidates to improve their clarity.

The remaining anchor candidates were included in a questionnaire instrument that was administered to 44 field experienced Army Warrant Officer aviators. These subjects were either current field aviators or students in the Warrant Officer Senior course (WOSC) at Fort Rucker. The subjects, who were mostly CW3 and CW4 ranks, were distributed across the four missions as follows: Aeroscout 20%, Attack 27%, Utility 23%, Cargo 30%. The anchor development questionnaire was adapted from the methodology used by Fleishman (1972-1973). Subjects assigned a value from 1-7 to each candidate corresponding to the amount of the given aptitude required to perform that task. Conceptual definitions were provided for each aptitude. The mean and standard deviation for each of the 288 anchor candidates were calculated and an attempt was made to select three anchors for each aptitude: one high, one low and one medium. In a few cases (6 of the 30 aptitudes) it wasn't possible to develop three anchors because the mean values clustered toward one end of the seven point scale, so only two were created. For each aptitude, the criterion was to select anchors that had small standard deviations, preferably 1.5 or less. The anchors were selected judgmentally to obtain the highest and lowest mean ratings having small standard deviations and also the rating closest to midscale (4.0) having a small standard deviation.

Rating Scale Evaluation. The rating scale approach to aptitude assessment was then evaluated by having Army aviators estimate the aptitude requirements of the four helicopter missions using the rating questionnaires with the aviation anchors served as points of reference on the seven point aptitude scales. The questionnaires were administered to experienced unit aviators (mimimum total hours 700, minimum hours in mission 200) at Fort Campbell, Kentucky: Hunter Army Airfield, Georgia and to a few combat and combat support unit aviators at Fort Rucker who had recently been gained from field assignments. A total ot 73 warrant officer aviators were sampled 19 Aeroscout, 19 Attack, 17 Cargo and 18 Utility.

Results

Inter-rates reliability. To estimate the inter-rater reliability of rating scales the aeroscout mission was chosen for detailed analysis. The data from the nineteen aeroscout aviators was factor analyzed across the thirty aptitudes. Analysis runs were conducted investigating the possibility of uncovering one through six factors in the data. However, if the interrater agreement fo the scales is high a single factor should account for the data. The results of the two factor analysis for the nineteen subjects are shown in Table 1. As can be seen from the table the

Table 1
Two Factor Loadings of Aeroscout Data

Factor 1	.63	.60	.60	.68	.83	.47	.69	.54	.82	.74
Factor 2	-.47	.38	.13	.03	-.16	.17	.06	.36	-.13	-.22
Factor 1	.84	.67	.84	.66	.49	.69	.16	.57	.70	
Factor 2	.19	-.07	.17	-.43	.30	-.39	.31	-.19	-.08	

data can be captured pretty well by a single factor. Fifteen of the nineteen subjects loaded on the first factor at over .5. No loadings on the second factor were over .5 and any second factor loadings between .4 and .5 were negative. Statistically 93% of the variance in the data could be attributed to factor one. This finding of a single facter indicates that most of the aviaters were performing the task in a similar manner and the inter-rater agreement of the rating scales was high.

Discrimination Among Aptitudes. To determine whether or not the rating scales were able to discriminate among the thirty aptitudes an analysis of variance was conducted on the data from the fifteen aeroscout aviators who loaded greater than .50 on factor 1 above. The results of this analysis showed that the rating scales were able to discriminate among the aptitudes (F. 29,14=13.6, $p < .01$). Given the successful analysis of variance, a Newman-Keules test was conducted to uncover any trends in the mean scores among the thirty aptitudes. The results of this analysis indicated that the aptitude ratings tended to fall statistically into three categories: primary requirements, secondary requirements, and incidental or low requirements. The aptitudes that were classed as primary or low requirements are given in Table 2. The remaining twenty aptitudes fell into the class of secondary requirements.

Table 2

High and Low Aeroscout Aptitude Requirements

Primary Requirements	Incidental Requirements
stamina	written expression
stress tolerance	visualization
time sharing	number facility
divided attention	static strength
perceptual speed	finger dexterity

Mission/Job Discrimination. To determine if the rating scale methodology was able to discriminate among the aptitudes required for the four different helicopter missions a two-way analysis of variance was conducted. The results of this analysis showed a statistically significant (F 29,001=30.07 p <.01) main effect of aptitude, again indicating that the rating scales were able to show differences among the aptitudes. However, both the main effect of mission and the mission helicopter interaction did not reach statistical significance (F 3,69=2.21 p=.095 and F 87,001=1.44, p=.76 respectively). Taken together these latter two findings indicate that the rating scales were not able to uncover any differences in the aptitudes required to fly the four different helicopter missions.

Discussion

The results discussed above showed that the rating scale methodology suceeeded in two of the three properties that were investigated. The methodology showed acceptable inter-rater reliabilities and was able to successfully discriminate among the levels of required aptitudes. Thus, it appears that the approach may be useful in analyzing Army jobs.

However, in this case the rating scale methodology was unable to distinguish among the aptitudes required to fly the four different missions. This finding is probably not surprising since the four different jobs are pretty similar. But it does indicate that the approach has limitations, and further research should be conducted to determine how much jobs should deffer before the rating scales will uncover different aptitude requirements.

References

Fleishman, E. A. On the relation between abilities and human performance. American Psychologist, 1972, 27, 1017-1032.

Fleishman, E. A. Toward a taxonomy of human performance. American Psychologist, 1975, 30, 1127-1149.

Kerwin, W. T., Blanchard, G. S., Atzinger, E. M., and Topper, P. E. Man-Machine Interface: A Growing Crisis (Among Top Problem Areas, Discussion Paper No. 2) US Army Material Systems Analysis Activity, Aberdeen Proving Ground, MD, 1980.

Rossmeissl, P. G. Kostyla, S. J., and Baker, J. Human Aptitude Assessment Techniques for System Designers. Proceedings of the 23rd Annual Conference of the Military Testing Association (Vol II), Arlington, VA 1981.

US General Accounting Office. Effectiveness of US Forces can be increased through improved weapon system design, PSAD 81-17. Washington, DC, Author, 1981.

Waters, B. K., Eitelberg, M. J., and Laurence, J. H. Military and civilian test score trends (1950-1980). Proceedings of the 23rd-Annual Conference of the Military Testing Association (Vol. II), Arlington, VA, 1981.

Organizations and Roles

USAREC is the proponent and developing agency for the Joint Optical Information Network (JOIN) System. The U. S. Army Research Institute for the Behavioral and Social Sciences (ARI) is providing research and technical advisory services to USAREC for the JOIN project. In FY79, the Navy Personnel Research and Development Center (NPRDC) began developing the Navy Personnel Accessioning System (NPAS), a system similar to JOIN, for the Navy Recruiting Command (NRC). Severe budget cuts suffered by NRC during FY81 resulted in a cancellation of funding for the NPAS Project for FY82 and the outyears. At this point, ARI proposed that the NPAS research team work on the JOIN System under ARI funding. An inter-laboratory agreement between ARI and NPRDC provides for a three-year effort beginning in FY82.

MAJOR FEATURES

Sales Presentation

Managers at USAREC Headquarters believe that the most critical area in the entire recruiting process is the sales presentation. As a result, the development of a sales presentation capability for the recruiter and the guidance counselor has assumed top priority.

Prior to the advent of JOIN, Army recruiters did have access to a Fairchild projector system using video cassettes to assist them in making sales presentations. However, current videodisc technology offers a far superior presentation. Moreover, the old system, with the numerous video cassettes, required considerably more storage space than does a videodisc player and a few, large-capacity videodiscs. This difference is important because physical space is at a premium in many recruiting stations.

The JOIN System will assist recruiters in presenting a realistic picture of the Army. This accurate and consistent information structures the development of realistic expectations on the part of the applicant. These realistic expectations, in turn, should have two direct benefits. First, both the incidence of recruiting malpractice charges and the attendant costs should decrease. Malpractice can result from a recruiter leaving a person with an impression of Army life which is found to be false after enlistment. A recent Government Accounting Office (GAO) report found that the most frequent cause of Army recruiter malpractice was misleading applicants about service conditions and benefits. The standardized videodisc presentation should drastically diminish this problem. The second probable benefit of realistic expectations is a reduction in premature attrition, a significant portion of which appears related to a discrepancy between pre-enlistment expectations and post-enlistment experiences. The JOIN System will allow the recruiter or guidance

AD P000902

RESEARCH AND DEVELOPMENT FOR THE JOIN SYSTEM

William A. Sands
Navy Personnel Research and Development Center
San Diego, CA 92152

Paul A. Gade
U.S. Army Research Institute for the Behavioral and Social Sciences
Alexandria, VA 22333

James D. Bryan
U.S. Army Recruiting Command Headquarters
Ft. Sheridan, IL 60037

INTRODUCTION

Recruiting Mission

The U.S. Army Recruiting Command (USAREC) has been tasked by the Department of the Army to enlist sufficient numbers of qualified young men and women to sustain desired force levels. The success or failure of "Manning the Force" depends directly on the Recruiting Command's "fighting forces;" i.e., the more than 5000 field recruiters located in more than 2200 recruiting stations across the country and overseas. Instead of being armed with rifles, these soldiers employ recruiting tools such as lead lists composed of seniors in a high school, combined with a high-powered national advertising campaign, to aid them in achieving their mission.

Market and Product

The target market for this recruiting effort consists of bright young men and women between the ages of 17 and 21. Recruiting from this market is required because modern weapon systems involve the latest available technology. It takes high quality personnel to use systems involving sophisticated computer and laser equipment. Unfortunately, this segment of the population is decreasing and forecasts indicate that this trend will continue throughout the decade. The increasing demand for high quality personnel, coupled with a decreasing supply, means that the Army recruiter will face stiff competition for high school diploma graduates, from the other military services and from colleges and universities which are facing declining student enrollments.

The product which the recruiter must sell to be successful in accomplishing the mission is a commitment, in the form of an enlistment. Ideally, in addition to "making mission," the recruiter will enlist many young people who will find that the Army offers numerous opportunities and benefits, and will choose to make the Army a career.

counselor to show an applicant a wide range of video segments, with audio commentary, to illustrate Army enlistment options, benefits, and training opportunities.

Another feature in the sales presentation is the collection of information on the applicant's stated interests and needs. This information is used by the recruiter to determine which factors will influence the applicant's enlistment decision, and thereby tailor the sales presentation to focus on these areas in order to obtain an enlistment commitment.

Person-Job Matching

Early in the recruiting process, the recruiter needs to know the likelihood that an applicant will achieve a qualifying score on the Armed Services Vocational Aptitude Battery (ASVAB). An accurate estimate of this likelihood of qualifying has a number of benefits. It enables the recruiter to avoid wasting valuable time with applicants who will not be eligible for enlistment. It reduces unnecessary costs for round trip transportation between the recruiting station and the test site. Finally, an applicant who spends the better part of a day attempting to enlist in the Army and finds that he or she is rejected, returns to the civilian community with a negative attitude regarding the Army. Communication of this negative attitude to friends will make the Army recruiter's job even more difficult in the future.

It is difficult to sell an individual on the idea of enlisting without providing some idea of the type of work that he or she will be doing. This is especially true for bright, high school diploma graduates. Under present USAREC policy, the job of the recruiter is to "sell" the Army, not a particular job or Military Occupational Specialty (MOS). Matching the individual with a specific MOS training slot is the job of the guidance counselor at the Military Enlistment Processing Station (MEPS). The recruiter can, however, discuss fourteen occupational clusters (e.g., electronics and communications) and show video segments of these MOS clusters on the JOIN System. The guidance counselor at the MEPS location will also have these video segments available, plus segments on the individual MOSs.

Management Support

USAREC is made up of a Headquarters, Five Region Recruiting Commands (RRC), a Recruiting Support Center (RSC), 56 District Recruiting Commands (DRC), 257 Recruiting Areas (RA), and over 2200 Recruiting Stations (RS). Like any large organization with an emphasis on sales productivity, USAREC is critically dependent upon an effective Management Information System (MIS). This system must insure that the right information is available in the right location, at the right time, so that decisions can be made with a minimal amount of guesswork. The effective communication of information within USAREC is a complex problem, involving over 2600 sites, many of which do not have access to AUTOVON or

AUTODIN. This existing situation creates real management problems. For example, if Headquarters decides to limit enlistments to high school diploma graduates for a specified period of time, the policy message may take days to reach all the recruiting stations. In the meantime, some Army recruiters, unaware of the policy change, will have invested considerable time in some applicants who do not have high school diplomas. Aside from the obvious waste of valuable recruiter time, this situation is quite likely to create negative feelings on the part of those applicants affected. These negative feelings tarnish the image of the individual Army recruiter as a professional, and the image of the organization represented. This, in turn, will have negative consequences for future Army recruiting.

Management of a large, distributed, information-dependent organization like USAREC requires the use of automated data processing equipment. Historically, business and government organizations employed centralized information processing, due to the high cost of mainframe computers. Recently, however, there has been a substantial decline in the cost of computing power and a proliferation of microcomputers, prompting decision-makers to re-evaluate the various means of satisfying their information processing requirements. The trend towards decentralization, wherein computing power is located where the work takes place, has been termed "distributed computing." The JOIN System is a good example of this trend, and, when augmented by an electronic mail service, will provide real-time management information throughout the Command.

At present, procedures for collecting and recording information on an applicant are manual and highly labor-intensive. For a single applicant, over 35 separate forms may have to be completed at various points in the enlistment process. Much of the information on these forms is redundant, resulting in multiple data entry tasks, unnecessary clerical time, and numerous administrative errors. The JOIN System, in conjunction with the Army Recruiting Accessions Data System (ARADS), will effectively solve this problem by capturing data at a single location and transmitting these data over communication links to other sites where they can be used to produce the required forms and management reports. USAREC is also examining all the various pre-printed enlistment forms to determine if some can be eliminated and to evaluate the possibility of revising the format of the remaining ones to facilitate their generation by computer. In addition, the JOIN System will provide recruiting personnel with a word processing capability. This will enable them to generate correspondence to prospects and to produce the numerous management reports necessary for production monitoring.

Personnel Training

Providing recruiting personnel with the training necessary to keep them abreast of changes in recruiting procedures and to maintain their skills and knowledge is a difficult problem for

the same reason that USAREC has communication problems; i.e., the geographical dispersion of over 5000 recruiting personnel in over 2200 locations across the country and overseas. When a new enlistment option is introduced or when substantial changes occur in some benefit, considerable recruiting time and expense is invested in some form of centralized training. Even when changes can be communicated entirely in writing, thereby avoiding centralized training, the written word is not always the most effective training medium.

The JOIN System will, to a large extent, solve this problem. The microcomputer and videodisc player combination provides an ideal vehicle for on-site training, using interactive, Computer Assisted Instruction (CAI) techniques. Moreover, on-site training can be accomplished at the recruiter's convenience, minimizing the disruption of normal recruiting activities.

RESEARCH AND DEVELOPMENT BY NPRDC

Aptitude Screening

At present, recruiters from the Army and the other military services are administering the Enlistment Screening Test (EST) to assess the likelihood that an applicant will achieve a qualifying score on the Armed Services Vocational Aptitude Battery (ASVAB). The EST is a conventionally-administered, paper-and-pencil test and, hence, suffers from a number of serious shortcomings: (1) excessive administration time; (2) relatively poor measurement precision at the extremes of the ability distribution; (3) susceptibility to test compromise; (4) cumbersome scoring and interpretation; (5) expensive and time-consuming replacement; and (6) limitations on the types of abilities which can be measured. Psychometric developments in the area of item response theory, coupled with technological advances and cost reductions in microcomputers, have provided an opportunity to address these shortcomings using Computerized Adaptive Testing (CAT).

The Computerized Adaptive Screening Test (CAST) has been developed by NPRDC and will be administered, scored and interpreted by the JOIN System. Item banks for Word Knowledge and Arithmetic Reasoning have been developed and the test items have been calibrated. Current research in this area involves field testing the instrument and administrative instructions on Army applicants and the development and evaluation of a prediction system for estimating an individual's Armed Forces Qualification Test (AFQT) score on the ASVAB. Plans call for implementation of the CAST on the JOIN System in early CY83, in conjunction with full implementation of the system nationwide.

Vocational Guidance

Work has been initiated to design and develop a comprehensive vocational guidance capability for the JOIN System. An extensive literature review on computer-based vocational guidance systems

is underway. The immediate purpose of this vocational guidance is to produce a well-informed applicant for enlistment. This will involve an exploration of values and career choices and the measurement of vocational interests. The long-range objective is to create realistic expectations and to facilitate person-job matches that are rewarding to the individual, while meeting the needs of the Army.

Personnel Assignment

At least for the forseeable future, actual personnel assignments will continue to be made by guidance counselors at the MEPS, using the Automated Recruit Quota System (REQUEST). However, it may be possible to provide the applicant and/or the Army recruiter with an estimate of the chances that the applicant will be offered one or more specific MOSs by the REQUEST System, given the MEPS arrival date and the date of availability for training. This information could be used to focus an applicant's attention on likely assignment possibilities, reducing indecision and hesitancy during the subsequent interview with the guidance counselor and facilitating a smooth, efficient enlistment.

Forms Generation

To date, the research and development in this area has been focused on the automation of the Application for Enlistment (DD Form 1966). NPRDC has developed a stand-alone system of interactive computer program modules, which automates the entire eight pages of the DD Form 1966. A central concern throughout the software development, test, and evaluation phases has been ease of use; i.e., the software must remain "user-friendly." Extensive documentation has been prepared for the system including a User's Manual and a Program Maintenance Manual. The software and supporting documentation have been submitted to ARI and, subsequently, to USAREC for conversion and integration into the operational JOIN System.

SUMMARY

The JOIN System combines state-of-the-art technology in several fields into a unique, powerful, computer-based, audiovisual, communications, and data management system that will benefit both Army applicants and recruiting personnel at all levels of USAREC.

Above and beyond the myriad contributions that the JOIN System will make towards meeting USAREC's mission of "Manning the Force," the potential uses of the system are not fully realized at present. Recruiting personnel, upon understanding the power which this system provides them, will discover new and creative applications which will enable them to enlist people into the Army better, faster, smarter, and with absolute integrity.

Geometric Radar Symbology: Static Cathode Ray Tube Testing

Gary G. Sarli and Richard J. Carter
US Army Research Institute for the Behavioral and Social Sciences (ARI),
Fort Bliss Field Unit, Ft Bliss, Texas

Many air defense (AD) systems use geometric symbols to indicate aircraft on system displays and different shapes to encode friend-or-foe information. The purpose of ARI's AD symbology research was to identify sets of geometric symbols associated with high discriminability and quick response times.

It was felt that AD personnel would respond fastest and most accurately to symbols with stereotyped meanings. Phase 1 of this research (Carter, in press) identified nine such symbols:

FRIEND: Circle (○), 5-Pointed Star (☆), Heart (♡), Flag (⚑)
HOSTILE: Swastika (卍), Collapsed Box (⌧), The letter "X" (X)
UNKNOWN: Question Mark (?), 6-Sided "U" (U)

Phase 2, using paper displays, tested these symbols in sets of three (1 of each type) and five (2 friend, 2 hostile, and 1 unknown) symbols. The three-symbol set with the quickest response time (RT) was Star-Box-U; the one with the least errors was Heart-Box-U. RTs for the five-symbol sets were not significantly different, but set Heart-Flag-Swastika-Box-Question Mark had the least errors. (Carter, 1980)

This paper deals with Phase 3, in which the symbol sets were presented upon a cathode ray tube (CRT) display. Symbol shape was the independent variable; RT and errors were the dependent variables. The hypotheses were that some sets would have lower RTs and errors, and that the results of Phases 2 and 3 would agree.

Experiments 1 and 2: Three-Symbol Sets

Method

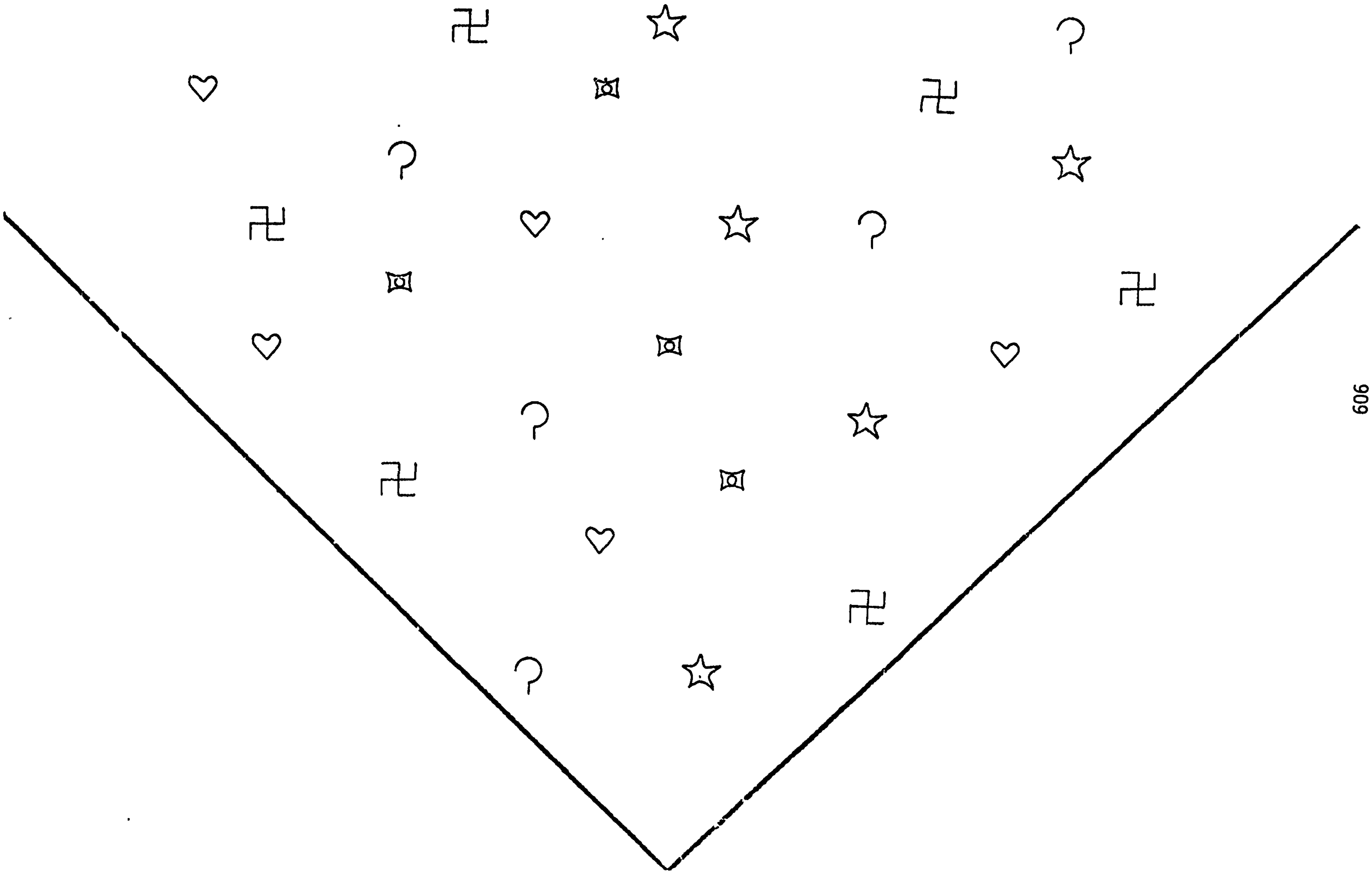
Subjects. For each experiment, 30 subjects were drawn from a pool of enlisted personnel (grades E2-E6) in AD console operator Military Occupational Specialties (16C, E, H, J) at the US Army Air Defense Center at Fort Bliss, TX.

Apparatus. A PATRIOT Tactical Operations Simulator/Trainer (TOS/T) provided a 36.2 cm diameter, round CRT, as well as necessary switches and controls, including an isometric joystick. A FORTRAN program was used to generate 64mm-wide symbols on the CRT and to record the data.

Procedure. A 90° top oriented search sector was plotted on the CRT (Figure 1). Each soldier, seated at the TOS/T console, was given 2 practice

The authors wish to thank CPT David Smull, SP6 David Arvieux, Jerry Price, and John Davis for their valuable assistance.

Figure 1
Example of CRT Display of Test Symbols



and 8 test sets (see Table 1) which each consisted of 1 friend, 1 hostile, and 1 unknown. On each set, the soldier was shown 27 scenes; each scene contained 24 symbols (7 to 9 of each type). The 27 scenes were divided into 3 groups of 9 scenes; the soldier was told to locate and "hook" either the friend, hostile, or unknown symbols in each group. The soldier hooked each symbol by using the joystick to superimpose a plus-sign-shaped cursor upon the symbol and then pressing the "hook" button. Feedback was provided by causing a PATRIOT "hold fire" modifier to appear around the hooked symbol.

Results

Experiment 1. The RTs for the eight symbol sets were significantly different: $F(7,203) = 2.391$, $P=.0225$. A Newman-Keuls' test showed that the soldiers hooked the symbols in the Star-Box-Question Mark and Heart-Swastika-Question Mark sets significantly faster than the two slowest sets. The number of errors made on the symbol sets were also significantly different, $\chi^2(7)=49.37$, $p<.001$, as were the numbers of errors made on the two hostile symbols, $\chi^2(1)=7.13$, $p < .001$. Symbol set Star-Box-Question Mark had the fewest errors among sets and the Collapsed Boxes had fewer errors than the Swastikas.

Experiment 2. The RTs for the symbol sets differed significantly: $F(7,203)=2.283$, $p=.029$; set Circle-Box-Question Mark was significantly faster than the slowest sets. The times for the friend symbols were also different, $F(3,87)=5.993$, $p=.0012$, with the Circle being significantly faster than the Flag. The number of errors were significantly different for sets ($\chi^2(7)=14.87$, $p < .025$), friends ($\chi^2(3)=17.16$, $p< .001$), and unknowns ($\chi^2(1)=7.13$, $p < .01$). Set Circle-Box-Question Mark, friend symbol Heart, and unknown symbol Question Mark had the fewest errors in their respective groups (see Table 1).

Experiments 3 and 4: Five-Symbol Sets

Method

Subjects. For each experiment, 24 subjects were selected in the same way as subjects in Experiments 1 and 2.

Apparatus. The apparatus was the same as that used in Experiments 1 and 2.

Procedure. Each soldier was given 1 practice and 6 test sets (see Table 2) which each consisted of 2 friends, 2 hostiles, and 1 unknown. On each set, the soldier was shown 50 scenes; each scene contained 25 symbols (4-6 of each shape). The 50 scenes were divided into 5 groups of 10 scenes each. The rest of the procedure was the same as in Experiments 1 and 2.

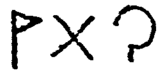
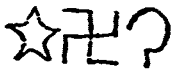
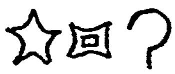
Results

Experiment 3. The RTs were significantly different only among friend symbols: $F(2,46)=6.707$, $p=.0031$. A Newman-Keuls' test revealed that the Hearts were significantly faster than the other friend symbols. The number of errors were significantly different for sets ($\chi^2(5)=9.51$, $p < .05$), and for friend ($\chi^2(2)=16.79$, $p < .01$) and hostile ($\chi^2(2)=12.52$, $p < .01$) symbols. Set Star-Heart-Box-X-Question Mark had the lowest error rate among sets. Hearts and

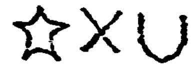
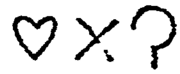
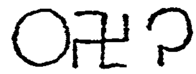
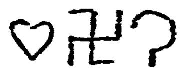
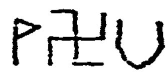
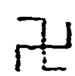
Table 1

Mean Response Times and Error Rates for Experiments 1 and 2

EXPERIMENT 1

Sets	RT	ERRORS
	(PRACTICE)	
	(PRACTICE)	
	322.07	2.80
	326.13	1.63
	308.15	.90
	314.36	1.27
	308.15	.97
	319.02	1.73
	315.71	1.72
	327.66	1.27
Friends		
	109.69	.55
	109.14	.42
Hostiles		
	102.31	.80
	103.31	.44
Unknowns		
	104.95	.42
	105.96	.45

EXPERIMENT 2

Sets	RT	ERRORS
	(PRACTICE)	
	(PRACTICE)	
	299.02	.83
	293.17	.34
	308.54	1.03
	302.97	.37
	312.81	.66
	309.12	.94
	311.95	.70
	302.77	.83
Friends		
	99.54	.19
	105.29	.60
	105.08	.10
	109.18	.42
Hostiles		
	96.30	.16
	96.29	.13
Unknowns		
	102.56	.18
	104.26	.37

Swastikas had the lowest error rates among friends and hostiles, respectively.

Experiment 4. Because only one unknown symbol was used, only sets, friend symbols, and hostile symbols were compared. The RTs for the sets were not significantly different, but the times for the friend ($F(2,44)=30.369$, $p < .00005$) and the hostile symbols ($F(2,44)=7.510$, $p=.0019$) were Newman-Keuls' tests revealed that the Circle was significantly faster than the other friends and that the X was significantly faster than the Collapsed Boxes. The errors were significantly different for the sets ($\chi^2(5)=19.44$, $p < .01$) and the friend symbols ($\chi^2(2)=55.64$, $p < .01$). Set Circle-Star-Box-X-Question Mark and friend symbol Circle had the lowest error rates for their groups (see Table 2)

Discussion

The hypotheses that the symbols and symbol sets would have significantly different time and error rates was generally supported. Note that response time differences were found among 3-symbol sets but not among 5-symbol sets. Inasmuch as the total number of symbols displayed at one time were comparable across the experiments, it is assumed that the effect of any one symbol on response time decreases as the diversity of symbols displayed increases (Earl, in press).

The hypothesis that the same sets and symbols would be identified as best in Phases 2 and 3 was not supported. This may be due to slight differences in the symbol shapes as displayed on the CRT versus on paper, the fact that CRT symbols were white-on-black instead of black-on-white, or differences in hooking procedures and practice sets. In general, the subjects in Phase 2 performed better on the Flag and six-sided "U" but worse on the "X" and Question Mark than the subjects in Phase 3. Clearly, paper simulations of CRT screens must be interpreted with great care.

References

Carter, Richard J. Paper-and-pencil testing of geometric radar symbols. Proceedings of the Military Testing Association, 1980, 22(1), CAR2-1 to CAR2-10.

Carter, Richard J. An investigation of geometric radar shapes for stereotyping (ARI Technical Report). Alexandria, Virginia: US Army Research Institute for the Behavioral and Social Sciences, in press.

Earl, William K. Learning and recognition of US, Soviet, and pictorial military symbology. Alexandria, Virginia: US Army Research Institute for the Behavioral and Social Sciences, in press.

Table 2

Mean Response Times and Error Rates for Experiments 3 and 4

EXPERIMENT 3

Sets	RT	ERRORS
○ ⚑ 卍 X ∪	(PRACTICE)	
○ ⚑ ⌧ X ∪	(PRACTICE)	
☆ ♡ 卍 ⌧ ?	498.67	2.46
☆ ♡ 卍 ⌧ ∪	493.49	2.83
☆ ♡ 卍 X ?	473.24	2.71
☆ ♡ ⌧ X ?	479.26	1.71
☆ ⚑ 卍 ⌧ ?	502.55	2.54
♡ ⚑ 卍 ⌧ ?	484.99	2.00
Friends		
☆	107.23	.80
♡	100.32	.40
⚑	105.46	.52
Hostiles		
卍	93.91	.30
⌧	94.74	.60
X	89.73	.40
Unknowns		
?	96.12	.31
∪	96.92	.38

EXPERIMENT 4

Sets	RT	ERRORS
○ ⚑ 卍 ⌧ ∪	(PRACTICE)	
☆ ⚑ 卍 X ∪	(PRACTICE)	
♡ ⚑ ⌧ X ∪	(PRACTICE)	
○ ☆ 卍 ⌧ ?	501.58	2.65
○ ☆ 卍 X ?	474.19	2.00
○ ☆ ⌧ X ?	477.46	1.57
○ ♡ 卍 ⌧ ?	487.12	1.61
○ ♡ 卍 X ?	491.77	1.57
○ ♡ ⌧ X ?	480.04	2.35
Friends		
○	94.51	.17
☆	105.07	.68
♡	108.73	.86
Hostiles		
卍	94.01	.33
⌧	97.44	.30
X	91.34	.39
Unknown		
?	94.90	.33

AD P000904

"WHERE ANGELS FEAR TO TREAD--
THE IMPACT OF CBI UPON INSTRUCTORS"

Worth Scanland, Ph.D.
Director, Research, Development,
Test, and Evaluation
Chief of Naval Education and Training
Pensacola, Florida

ABSTRACT

Computer Based Instruction (CBI) has been around in one of its several forms or another for the last two decades. Both in research settings and in schoolhouses it has consistantly shown to be able to deliver high quality instruction to almost every level of student in most subject matter areas, provided that the courseware is written by skilled practitioners. Yet the world of education and training continues, in the face of mountains of evidence to the contrary, to stick with the belief that, "CBI is a valuable adjunct among the many other instructional aids, such as audiovisual, available to the instructor, but it will never take the place of that instructor."

COMPUTER BASED INSTRUCTION

It is the purpose of this paper to clear the air of the very erroneous and harmful notion that CBI can not and will not replace many instructors. To believe, or be deceived into believing, the contrary is doing both the instructors and CBI a disservice, and may if not countered, lead us to the costly error of failing to grasp the full benefits which CBI is about to be able to deliver as the costs of both hardware and software continue to sharply decline.

Several weeks ago I attended a seminar conducted at the loc l community college on the subject of computers in education. The audience was made up principally of youngish men and women engaged, in one form or another, in public education and military training. They were curious about computers and the role computer based instruction might play in their world of work, and I was there to hear what the world of higher education had to say about that subject. The speaker, he said, was a doctoral candidate at a university well known for its curriculum in educational technology (in fact I took my graduate degrees there and regard its teachings most highly). We were all appropriately impressed with his credentials to speak on the topic at hand.

The seminar opened with the leader saying something such as the following, and my quotation, if not verbatum, is at least accurate in its sense. "Let me set the record straight for us right here at the beginning," he said. "A computer will never (and he underlined with his voice the timelessness of his prediction)...the computer will never take the place of a teacher. Let me ask you a question," he went on. "Have you ever tried to imagine a computer drying the tears of a little six year old girl?" I

stood up and left the room at that point, unable to listen to more of that sort of pablum about the relationships between teachers and computer based instruction, yet realizing as I did so that nowhere have I really heard the subject openly and frankly addressed. I think the time has arrived for us instructional technologists to cease apologizing for advances in our technologies, and begin to face the facts, disconcerting to some as they may be. Let us, therefore, "Go where angels fear to tread", and, to thoroughly mix our metaphors, "Grab the bull by the horns!"

Prior to about 1400 A.D., when either the Dutchman Coster or the German Gutenberg invented movable type (and you may take your pick), books were the exclusive communications medium, other than word of mouth, of the church hierarchy and royalty, and common man remained ever forbidden the truths of science and the wisdom of the arts. But with that revolutionary invention a whole new world of education was opened up to the entire populations of the world, and a major step forward toward science, medicine, geographic discovery, navigation, mathematics and social equality was born. Books were our primary medium of communications between the peoples of this earth until early in the 19th Centruy, when Samuel Morse invented the telegraph which allowed the written word to be transmitted all over the globe at the speed of light. And then Alexander Graham Bell brought forth the first telephone, allowing people of all ages and races and nationalities to exchange information by orally spoken words. One must grasp the historical significance of these events and their impact upon mankind in order to place the computer in its true perspective...invented in the middle of the 20th Century, the computer is as significant to mankind and his abilities for rapid and ubiquitous communications as were those inventions which preceeded it, and perhaps history will someday tell us, even more important. We are emerging into a "brave new world," as Aldous Huxley named it, where the computer will be as essential to the conduct of human affairs as has been the printed page, the telephone, the television, the automobile and the airplane, only far more so than we can, even today, imagine. It will become the constant and ever-present medium through which we exchange ideas, knowledge, scientific fact and the arts, and by which we will order our daily lives, businesses, production, economics, both public and private, <u>and</u> our education and training. If you don't believe that, you are as far behind the times as those who burned the early books because they were heretic.

As you are likely aware, the early computers which appeared on the scene about 1948 to 1952, were in fact simply very fast mathematical calculating machines. Although they could, in fact, only perform the basic functions of addition and subtraction, they could do so at such speed that other mathematical functions could be performed, or seem to be performed, as well. Fantastic as these developments were, they had been envisioned more than a century ago by a most astute lady named Ada Augusta Lovelace, who said and I quote from her writings, "--The Analytical Machine (her name for a computer) holds a position wholly its own; and the considerations it suggests are most interesting in their nature. In enabling mechanisms to combine together general symbols in successions of unlimited variety and extent, a uniting link is established between the operations of matter and the abstract mental processes of the most abstract branch of mathematical science. A new, a vast, and a powerful language is developed for the future use of analysis, in which

to wield its truths so that these may become of more speedy and accurate application for the purpose of mankind than the means hitherto in our possession have rendered possible." That, my friends, is an amazing statement, considering that it was pronounced before the Civil War took place.

It was through the progression of computers to the point where they could deal with other forms of communications than mathematics, principally the written word and other symbols conveying ideas, facts and concepts, that the real breakthrough occurred. Once this had been achieved the floodgate of computer capabilities had only to await the development of languages with which to communicate to the computer itself, and the engineering and technologies to produce smaller and faster operating systems. These advances have been the hallmark of computer technological advances over the decade of the seventies, and the advent of the personal and home computer is the obvious result.

At the risk of saying something which is fairly apparent even to the least enthusiastic computer observer, let me point out to you a very significant perspective. In Figure 1 we see a graph on which the absissa represents time in years, 1950 to 1990, and the ordinate represents the cost of computer capability, in whatever units you might select. What the curve says is that back in the 1950s any usable computer capability, let's say to solve complex calculus equations, cost millions of dollars (and, I might add, occupied hundreds of cubic feet of space.) As the years went by this curve took a definite turn downward, until by the year 1980 we could purchase handheld computers which had the capability to perform complex engineering equations with the touch of a button and a PDP-11/23 could predict the trajectory of a space ship to the moon and back, at a cost of less than one tenth that of a 1960 IBM 1500. Today one can buy at the local computer store a Timex-Sinclair 1000 for less than one hundred dollars, and can bring it home, plug it into any 120 volt 60 Hertz outlet and talk to it like a Dutch Uncle! If that doesn't impress you, let me say that I heard the Chief of Naval Research, Rear Admiral Kollmorgan, describe a computer which will be on deck by 1986 with the computing capacity of a PDP-11/23, all packaged on a wafer three inches in diameter. So inexpensive computer capability will very soon blanket the world just as have paperback books and the Sony "WALKMAN."

Now then, what has all this to do with computer based instruction, and the role of the teacher in that new world?

Computer assisted instruction, along with its companion piece, computer managed instruction, express the primary applications of the computer to instruction generically known as computer based instruction (hereinafter referred to as CBI). The advent of CBI followed the development of the early computers by about a decade. In the middle of the 1960s CBI had been reasonably well researched as to its efficacy and ability to perform the instructional role, especially in the provision of facts, concepts and ideas, and the management of practice and testing. Most CBI programs used the typical "programmed instruction" format, which is fairly easy to author, and suited the programming limitations then prevalent. By and large these experimental systems were sponsored and funded by the U.S. Department of Education (then H.E.W.), and resided in universities from Stanford on the West Coast to Florida State University in the East. There were numerous public school demonstration projects of CBI, these also

funded by the Government. In summary, it can be safely said that throughout the sixties and into the early seventies there were more than adequate demonstrations and evaluations of various forms of CBI to persuade those who kept open minds that this new form of instructional delivery could handle the job as well as the best of teachers, and better than the majority. Only in the affective domain is it not at its best, but fortunately little instruction, compared with the total, is in that category. However, an history of that period would reveal that when the support of the Federal Government was withdrawn (and quite properly), the states and communities were not financially capable of sustaining these experimental CBI systems, and they, by and large, folded for lack of support. The plain fact of the matter was that, regardless of the demonstrated utility of the CAI concept, the cost of both hardware and software was prohibitably high. Thus for the better part of a decade CAI, with some isolated exceptions kept alive by either a farsighted industry or a university, lay dormant. There was, fortunately, one form of computer based instruction which was, even then and still is, quite affordable, namely computer managed instruction, or CMI. In this form of CBI one computer can handle as many students simultaneously as is required. For example, the Navy's CMI system, supporting its technical training program, has a large general purpose computer located near Memphis, Tennessee. It currently has ten thousand students registered on the system at any one time, and manages their instruction without significant problems. This application of the computer to the instructional process does not bring the students into a face-to-face interaction with the computer program, as does CAI, and therefore demands much from the student in the way of self-teaching. Nevertheless, it is a very, very efficient way to manage individualized instruction, and has saved the Navy millions of dollars in training time.

But despite the success of CMI where it has been applied, it is far from the ideal way by which to take advantage of the computer's capability to deliver instruction, and therefore, while not as dead as the dodo bird, will soon be overtaken by CAI. How, considering the costs factor we have mentioned briefly as being the reason CAI has not proliferated, can we so confidently predict the future success of CAI? The reason is quite simple...it resides in the cost/capability vs time curve we showed in Figure 1. The fact is that the costs associated with large computer capability and high computing speeds have plummeted over the last few years, to the point where today anyone with a modest income can afford his own computer, and compared with the costs of a human teacher (in Navy training about $32,000 a year), a CAI computer is fast becoming very, very affordable.

The purpose in taking you down this path with me on the brief history of CBI has been to permit us all to address the issue before us--to wit, what is the role of the instructor in CBI--on a common understanding of just what that instructional methodology is all about.

There are seven primary instructional functions which any teaching system, be it a tutor, a classroom instructor or a machine, must address. These are, sort of in the order of their appearance on the stage, as it were, as follows:

- Information presentation

- Demonstration

- Drill and practice
- Evaluation
- Feedback
- Remediation, and
- Instructional management

Of these seven, CMI takes care of only the evaluation, feedback, to some extent the remediation, and the instructional management functions. CAI takes care of them all, with bells on. How well can the average, or even the best, human teacher get around to performing all these functions? If you know anything about the average classroom instruction you know that all may get some attention to a limited extent, but that normally most are provided in half measure, if that well. Fifteen years of research, experimenting and demonstration have conclusively proven that good CAI (and let me say that anything but good CAI is worse than no CAI)...that good CAI can handle all seven functions with ease, and do them very well indeed.

If this is the case for CAI, and it is affordable, even economical when compared with human instructors, where is the problem? The problem is that the introduction of CBI, as with the introduction of any new major change in the accustomed way people do things, is threatening and therefore to be resisted.

It is a fact that the introduction of new technologies does, in many instances, replace the people who heretofore performed the job now capable of being done by the new technology. The introduction of machine tools into the British world of production and manufacturing threw thousands of workers out of their jobs, as is the introduction of robotics into the automobile production lines today. But it is also an historical fact that machine tools increased the demand for British goods by orders of magnitude, and new roles for working people grew from that demand. The end result has always been an overall increase in gross national product and a better standard of living for our society. Automation of many manpower intensive functions in the society is going to demand changes in the roles of those impacted, and the solution is historically evident that those people must make adjustments, primarily through retraining and education.

So it will be, and even is today, with the introduction of automation into the world of education and training. My personal experience in public school education is limited pretty much to twenty years as a student, so I will leave the debate on this topic in that setting to others more qualified. But in the domain of technical training I am perfectly willing to take my stand that the automation of instruction will, among other good things, result in the replacement of human instructors with computer based instruction. Now that doesn't necessarily mean that schools are going to have to offer up some of their instructor billets: In the case of the current Navy initiative to implement widespread application of CBI the computers will take the place of instructors who would otherwise have to be hired to accommodate the planned increase in student load. But we should not let that sophism lull us into the belief that CBI will not, tomorrow or the day after tomorrow, take the place of instructors now on the job. Not all of them, mind you. I do not foresee that day. But some of them, to be sure. One way around the problem, shortsighted as it appears to us

all, is to assure and reassure the threatened instructors that CBI cannot replace them in the system, and that it will only be another adjunct among the many they are accustomed to which will, and I quote, "Make the instructors' task easier.") We must not make this mistake.

The student-to-instructor ratio in the Naval Education and Training Command at this time is ten to one, which includes among the instructors those in training, those on annual and sick leave, maintaining curricula, or otherwise not necessarily on the podium. If we are serious about taking advantage of the potential for instruction inherent in the computer we should look forward over the next five years to that ratio becoming more like twelve or thirteen to one. In other words, we should expect, even demand, at least an increase in instructional productivity on the order of twenty to twenty-five percent. If we do not achieve that, we are failing to take advantage of the opportunities made available to us through automation.

But there is a brighter side of this picture from the standpoint of those instructors who enjoy playing a role in training. For every instructor who must give up the podium there is a place, a major requirement, for people who can provide routine maintenance to computers, repair them when they fail, and most of all, for people who can author well designed and developed instructional programs to be placed on those computers. Such changes in roles requires re-training, of course, and it is our duty to plan for and implement such re-training programs for our institutions, starting right now. CBI is here...large scale CBI is just around the corner. The time to plan for the future is today, and I urge each of you who has a responsible position in a training management situation to start looking out ahead. Change is upon us, and CBI is a role changer. Let's be prepared to meet the future instead of shying away from its challenge.

The view expressed in the above paper is the personal and professional view of the authors, and is not intended in any way to represent that of the commands by whom they are employed.

WORTH and DOROTHY SCANLAND

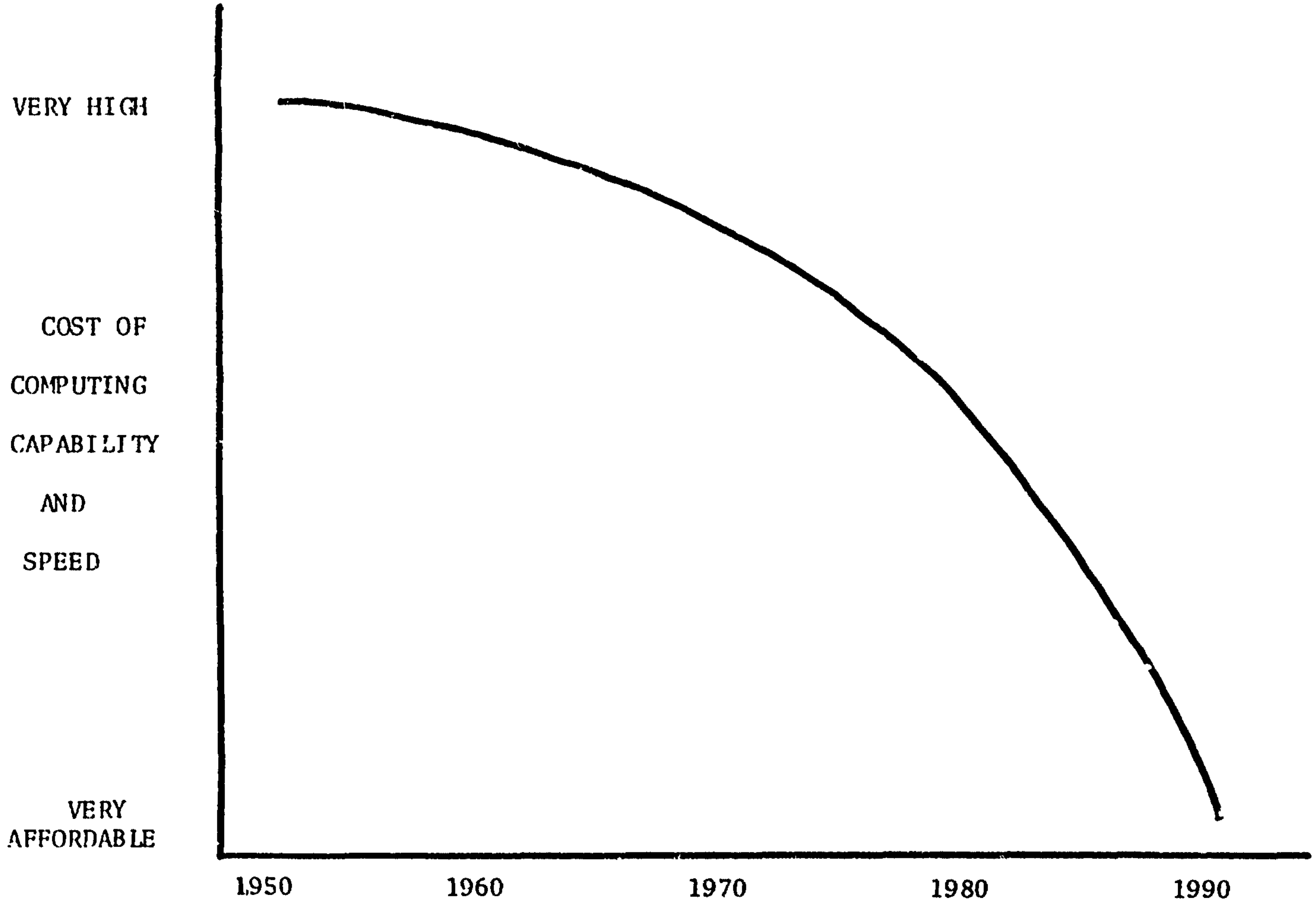


Figure 1.

THE TREND OF COMPUTER CAPABILITY VS TIME IN YEARS

Women in the Navy's Civilian Skilled Blue-Collar Workforce[1]

Amiel T. Sharon and John D. Kraft
U.S. Department of the Navy U.S. Office of Personnel Management

Although women have traditionally not been well represented in the skilled blue-collar workforce of the United States, their numbers have grown dramatically in recent years. The relatively high pay, affirmative action programs, the generally higher level of women's participation in the workforce, and the women's liberation movement have all contributed to their entry into the skilled trades and crafts in significant numbers. The U.S. Navy, which is among the largest employers of craftsmen in the nation, has also experienced a rapid rise in the number of women in virtually all of its skilled trades. The Navy currently employs over 50,000 blue-collar workers in trades such as welding, air conditioning and refrigeration, and electronics. These workers perform a vital function for national defense -- they overhaul and repair ships, aircraft, and complex weapons systems. They learn the skills of their trade in a grueling four-year apprenticeship, receiving both classroom instruction and on-the-job training.

Method of Survey and sample description

As part of an effort to develop and validate new procedures for the selection of apprentices to naval trades and to improve apprentice training programs, a comprehensive occupational survey of a representative sample of almost 5,000 skilled workers in 22 trades was conducted by the U.S. Office of Personnel Management and the Navy Department. The survey was accomplished through a structured questionnaire that sought information from job incumbents about educational and occupational background, methods of recruitment and selection to present job, perceived discomforts and hazards of the job, injuries sustained, relevance of classroom apprenticeship to the job, major job duties, physical demands of the job, tools and equipment used, and job tasks. The survey questionnaire was administered in group sessions at 18 different Naval installations.

The total sample was composed of 4,646 males and 197 females distributed among all of the 22 trades that were surveyed (a list of the trades is presented in the Appendix). Although no attempts were made to match the male and female subsamples, a coincidental (and fortuitous) finding was that the two groups had the same average length of time working as civilian in the Navy (7 years). This finding facilitated the comparison between the two subsamples because length of experience in the Navy might have been a confounding factor in the comparison of other variables.

The female subsample was found to have a higher average educational level than its male counterpart. Although there were slightly more women than men without high school degrees (17% women vs. 12% men), 54% of the women had some

[1] The opinions expressed are those of the authors and do not necessarily represent the positions of the Department of the Navy or the Office of Personnel Management.

college or a college degree in comparison to only 40 percent of the men. This finding may be related to the type of job that job incumbents held immediately before their current job. Twenty-one percent of the women as compared to only 12 percent of the men had prior jobs in a technical or professional area. Men, on the other hand, were more likely to have jobs as journeyworkers or apprentices in private industry (27% men vs. 11% women) or in the military (9% men vs. 2% women) prior to working for the Navy.

Recruitment and selection

By far, the most common way for an individual, whether male or female, to learn about Navy's apprenticeship program is through relatives or friends who work for the Navy. For more women than men, this was the major source of finding out about apprenticeship opportunities in the Navy (29% women vs. 22% men). Other recruitment methods were generally not effective with either sex. The reasons for wanting to become skilled journeyworkers were similar for men and women, with one exception. Pay and benefits led the list (28% for each group) and was followed in the female subsample by interest in the technical and physical aspects of the job (16% women vs. 17% men). Job security, however, was a much stronger motivating factor for men (24%) than for women (14%).

The finding that only 19 percent of the women as compared to 73 percent of the men were military veterans was not surprising. Since the civil service selection process for apprentices requires the awarding of veteran's preference points to qualified veterans, fewer women than men had the veteran's preference advantage in the selection process. Many women were hired into the apprenticeship program from white-collar jobs through merit promotion and upward mobility programs administered internally within the Naval installations at which the women were already employed in other capacities.

Major Job Duties

Survey respondents were asked to review a list of 27 major job duties typically performed by journeyworkers in the skilled trades and indicate whether these duties apply to their jobs. The percentage of men and women who perform each of the duties was fairly similar. Five duties, however, were found to differ by more than 10 percentage points between the two groups. For each of these duties the percentage of men who performed them exceeded that of women. The duties are:

- Teach or instruct trainees about job
- Lead or oversee the work of others
- Prepare written reports
- Give oral instructions
- Use or calculate fractions, decimals, or percentages

We can only speculate on the reasons why fewer women perform these duties. One key reason may be that a higher percentage of women in the survey sample were in apprentice positions (23%) than were men (9%). (The remainder were mostly in journeyworker level jobs.) Four of the five duties are associated with directing the work of others, the type of duties apprentices are not prepared or qualified to perform.

Discomforts, hazards, and injuries

The survey respondents were requested to indicate which of several types of physical discomforts and hazards they encountered on the job. Fewer discomforts were reported by women than men in every one of 13 categories of discomforts such as long periods of standing and frequent kneeling or stooping. The pattern of discomforts for both sexes, however, was similar. Working in a noisy environment, for example, was the most discomforting aspect of the jobs for both groups. Similar results were found with the hazards or dangers frequently encountered on the job. Fewer women than men reported that their jobs were hazardous in every one of 12 categories such as intensive heat that could lead to burns or work that could lead to muscle strain. Again, the pattern of perceived dangers was very similar for both groups, with exposure to contaminated air and exposure to excessive noise being cited most frequently.

The women fared better than the men in injuries that resulted in absence from work. Fifty-three percent of the women, as compared to 43 percent of the men, never incurred such injuries. Significantly fewer women reported receiving injuries to extremeties such as toe or finger; face, eye or ear injuries; back injuries; and hernias. Approximately the same percentage of men as women experienced scrapes or bruises, chemical burns, and damage to lungs from smoke.

Physical demands of the job

We recently became aware of the decision by the U. S. Army to restrict women from performing in many arduous enlisted jobs. One article in The Washington Post (1982) said that women would be barred from 76% of Army Military Occupation Specialties (MOSs) because of the new strength requirements.

While the report on these research findings has not yet been released by the Army, we did attend a medical briefing by Major Dennis Kowal, the research psychologist in charge of the research basis for these policy decisions. As a result of this briefing, it appears to us that these findings should not have impact on the ability or desirability of the Navy to hire, train, and utilize more women in civilian skilled blue-collar positions. We think that this is true for several reasons.

First, the legally mandated combat MOS automatic exclusion of a majority of positions for women was not reported in the papers. This policy, infact, already excludes women from most of the physically-demanding Army jobs.

Second, while many of the military enlisted jobs are classified in the same Dictionary of Occupational Titles (DOT) codes as the civilian jobs in the Navy, they are quite different in how they are carried out. In order to insure military preparedness, enlisted personnel are assigned to specific two or three person teams who work together on tasks; no extra persons are available to perform a particularly arduous task. This is not true for civilians; usually there are other persons around who can help perform a particularly demanding aspect of a task.

Third, the research literature does not report significant differences in industrial injury rates for men and women (Ballau and Buchman, 1978). An

exception to this literature was the research by Major Kowal on injuries during basic training. However, Kowal indicated that these injuries were directly related to the sustained nature of basic training and the resulting deterioration over a short period of time of body structure before it is rebuilt (Personal communication and Kowal, 1980).

Fourth, different laws affect the military and civilian personnel policies. The Civil Rights Act of 1964, as amended, and the Rehabilitation Act of 1973, as amended, do not apply to military personnel but do apply to Federal civilian personnel. As a result, the case law developments and regulatory agencies' regulations apply to Navy's policies in this area. The Federal government as a civilian employer cannot exclude women from a job because of so-called bona fide occupational qualifications, such as strength requirements, except in extremely rare circumstances. This includes such industrial hazards as the use of solvents and similar chemicals which might pose reproductive system hazards (Stillman, 1979). Even cases of lower back injury will not exclude persons from employment in most arduous jobs. Recently, a Federal judge citing Rehabilitation Act considerations ordered a construction company to hire and provide back pay to an apprentice carpenter who was denied employment because of a lower back condition. The basis of his decision was that the possibility of future injury did not constitute grounds for disqualifying the carpenter from the job (Medical Standards News, 1981).

Our own survey results show few significant differences between men and women in coping with the physical demands of the job. Using a taxonomy of occupationally-oriented Basic Body Efforts developed by researchers at the Naval Personnel Research and Development Center, survey respondents were asked to provide information about the single most muscularly demanding task of their job. No significant differences were found between men and women in the type of effort applied to the most demanding job task. One-half of each group reported that carrying objects while walking, such as carrying a motor to the shop for overhaul, and lifting objects without carrying them, such as lifting a box onto a shelf, were their physically most demanding tasks. Other types of effort were reported by relatively few workers. Significantly more women (49%) than men (35%) perform the most demanding task alone as opposed to being teamed together with other workers to exert the required force. Although most individuals in both groups indicated that they are able to perform the most demanding tasks of their job without problems, 18 percent of the women as compared to 10 percent of the men reported that the task sometimes exceeded their strength capabilities. More women than men (40% vs. 32%) reported that the task is difficult to perform because of the pounds of force exerted. Fewer women, however, reported problems as a result of the difficult grip (47% vs. 57%), cramped or restricted spaces which restrict body leverage or movement (28% vs. 49%), and the reach required to move or install an object (32% vs. 43%).

Conclusion

The results of the survey suggest that there are some differences and many similarities between men and women working in Naval skilled trades. The differences, which were highlighted in this paper, should not overshadow the preponderance of similarities between the two groups. Most of the evidence suggests that women in naval skilled trades are not unlike their male counterparts in the duties they perform and in the difficulties they encounter on the job.

623

from p. 622

Based on our limited survey findings, literature on injuries, and the present interest in increasing the number of women in underrepresentative civilian jobs, it may be that a larger proportion of the Navy's skilled blue-collar civilian work force will be women in the future.

REFERENCES

Ballau, R. I. & Buchan, R. M. Study shows that gender is not a major factor in accident etiology. Occupational Health and Safety. 1978, 12, 54-58.

Gloss, D. S. & Wardle, M. R. Ergonomics and the Working Woman. Occupational Health and Safety, 1979, 11, 20-23.

Kowal, D. M. Nature and causes of injuries in women resulting from an endurance training program. The American Jounal of Sports Medicine, 1980, 8, 265-268.

Medical Standard News. Court Backs DOL E.E. Black Order in Part. San Bernardino county, Author, Winter 1981.

Stillman, N. G. The law in conflict: accomodating equal employment and occupational health obligation. Journal of Occupational Medicine, 1979, 21, 599-605.

Washington Post. New tests for strength will bar most women from 76% of army jobs. Washington, D.C.: Author, August 31, 1982.

APPENDIX

NAVAL TRADES AND CRAFTS SURVEYED

Air-conditioning Equipment Mechanic

Aircraft Electrician

Aircraft Engine Mechanic

Aircraft Instrument Mechanic

Aircraft Mechanic

Boilermaker

Boiler Plant Operator

Carpenter

Electronic Mechanic

Electrician

Electroplater

Equipment Mechanic (formerly Marine Machinist)

Inside Machinist

Insulator

Painter

Pipefitter

Rigger

Sheetmetal Mechanic

Sheetmetal Mechanic (Aircraft)

Shipfitter

Shipwright

Welder

Transition Socialization Processes in the U.S. Marines*

James B. Shaw, Department of Psychology
Richard W. Woodman, Department of Management
Texas A&M University

A longitudinal study of socialization processes in U.S. Marines who were transfered from Camp Pendleton, California to Okinawa was begun in August 1981. This paper presents data relating to expectations concerning Okinawa that marines develop and the effect of expectations upon several job-relevant attitudes and behaviors.

Van Maanen and Schein (1979) have defined organizational socialization as "the process by which an individual acquires the social knowledge and skills necessary to assume an organizational role" (p.211). Feldman (1981) breaks the socialization process into three stages: (1) anticipatory socialization, (2) encounter, and (3) change and acquisition. It is this first stage of anticipatory socialization that is under scrutiny in the present study. Van Maanen (1976) defines the anticipatory stage as being concerned with "the degree to which an individual is prepared - prior to entry - to occupy organizational positions" (p.81). Feldman (1976) suggests that one important factor in determining this degree of preparation is the realism of information concerning the organization and job which the individual has prior to entry. Research (Ilgen & Seely, 1974; Wanous, 1973; and Zaharia & Baumeister, 1981) has shown that realistic job information provided prior to beginning work facilitates role adjustment. Much of this work has centered around the notion that information given prior to organizational entry results in the individual developing expectations about the job. These expectations (or lack thereof) affect the ability of the person to adjust to the new job situation.

In the present study, we were interested in the relationship between expectations and adjustment in transfers rather than initial organizational entry. Although Fisher, Wilkins and Eulberg (1982) point out that entry and transfer situations differ somewhat in the nature and extent of anticipatory socialization that would occur, they nevertheless suggest that accurate pre-transfer perceptions will be as important as pre-entry perceptions.

The present study we examined the role of expectations developed prior to transfer on later transfer adjustment. Specifically we investigated the relationship between expectations developed prior to transfer, their "realism," and later adjustment in the new job situation.

METHOD

Sample

The sample consisted of 91 marines stationed at Camp Pendleton, California who were scheduled for transfer to Okinawa in November 1981. Over 90% of this sample were assigned to 2nd Battalion, 7th Marine Regiment. For 33% of the sample, the move to Okinawa would be their first experience outside the U.S. Over 90% of the sample were E-4's or lower. Nine percent

*This research was funded by the Office of Naval Research, grant number N00014-81-K-0036, project NR 170-925.

of the sample were married. The sample was not selected to be representative of 2nd Battalion, but was instead weighted with younger, less experienced personnel.

Procedure

Preliminary interviews with personnel recently returned from Okinawa were held in August 1981. These interviews were used to develop and refine questionnaire items. In October 1981, 91 individuals about to transfer to Okinawa completed the survey instrument and were interviewed. Follow-up interviews were conducted with 79 of these marines in April 1982. The same survey instrument was completed. Minor wording modifications of some items were made for the Okinawa sample. For example, an item which for the Pendleton administration was worded "what do you expect Okinawa to be like ..." would be reworded "what has Okinawa been like ..." For each of these expectation items, responses from Pendleton and Okinawa were available. For each marine a difference score, Okinawa-Pendleton was computed for each item.

Measures

Data were collected using the survey instrument. Expectations of marines about living in Okinawa, their job in Okinawa, the natives in Okinawa, their standard of living while in Okinawa, drug and alcohol use in Okinawa, closeness among members of their unit while in Okinawa, officer strictness, rule enforcement, weather, entertainment opportunities, number of friends they would have, how much they would miss their family, and their overall difficulty of adjusting to Okinawa were collected as part of the survey. Five point rating scales were used.

Also included in the survey were items concerning various transfer adjustment related attitudes and behaviors. Specifically, marines were asked to respond to items concerning their intention to complete their enlistment, their intention to re-enlist, their overall satisfaction with the Marine Corps, their preference for being stationed in Okinawa or Pendleton, and the amount of time they felt it took them to adjust to being in Okinawa. Marines also indicated the number of times per week they "got really angry and told someone off," got in a fight, got drunk, used drugs other than alcohol, went on an unauthorized absence or were put in the brig.

RESULTS

Expectations

Subjects answered 13 questions designed to measure various aspects of their expectations about their new assignment. For both the Pendleton and Okinawa administrations data on response frequencies to each of the items as well as the mean response to each item were obtained. The results of t-tests conducted for each item between mean responses in the two administrations (Pendleton vs. Okinawa) are discussed below. The Pendleton sample refers to the ninety-one marines from whom data was collected at Camp Pendleton prior to their departure for Okinawa. The Okinawa sample refers to the seventy-nine individuals interviewed, for the second time, during their stay on Okinawa. In order to be included in the t-tests, a respondent had to complete the item in both data collections. Thus, the n for this t-test varies across items.

<u>Item: "How interesting living in Okinawa would be."</u> Almost 62% of the Pendleton sample expected living to be more interesting in Okinawa, although only a little over 20% actually later found living in Okinawa more interesting. In like manner, over 14% thought living there would be more

boring, but almost 70% found it more boring. The difference in mean responses between the Pendleton and Okinawa samples (3.54 versus 2.22) is statistically significant ($p < .001$).

<u>Item: "My job in Okinawa will be."</u> Almost 54% of the marines expected that their job would be more interesting in Okinawa, but only 38% found it so. However, significant numbers of respondents both expected (44%) and found (54%) the job to be about the same as they had experienced before. The difference in mean responses between Pendleton and Okinawa (3.68 versus 3.38) was not quite large enough to meet a criterion of statistical significance ($p = .053$).

<u>Item: "The natives on Okinawa will be."</u> Marines answered this question using a five-point scale that ranged from very hostile (1) to very friendly (5). The majority of respondents (60%) expected the Okinawans to be "indifferent" while almost 32% expected them to be friendly. Both the percentage of marines rating Okinawans as friendly (42%) and as hostile (14%) were higher than the expectations in each category. However, mean responses to this item in the Pendleton and Okinawa samples are virtually identical.

<u>Item: "Compared to here, my standard of living while in Okinawa will be."</u> Using a five-point answer format (ranging from 1 = much poorer to 5 = much better), a slight majority of marines expected their living standard to remain about the same. However, almost 47% of marines reported that their standard of living dropped (only 26% had expected this to happen). Some 15% of the Okinawa sample reported an improved standard of living, while 23% had expected improvement. Mean responses declined almost one-half point on the five point scale (from 2.98 to 2.53), and this drop is highly significant statistically ($p < .001$).

<u>Item: "In Okinawa, drug and alcohol use in my unit will."</u> Almost 41% of the Pendleton sample expected drug and alcohol use to decrease in Okinawa. Some 35% of the Okinawa sample reported a decrease. Over 38% of the marines expected an increase in drug use and a majority (almost 56%) reported such an increase. While mean responses to this item indicated a perception of increased drug and alcohol use in Okinawa, the difference between Pendleton and Okinawa responses is not statistically significant.

<u>Item: "While in Okinawa, my unit will be."</u> Marines responded to this item using a five point scale that varied from "much less close" (1) to "much closer" (5). A large majority (76%) expected greater closeness after their unit arrived in Okinawa, but a somewhat smaller percentage (although still a majority at 63%) reported greater closeness. Few marines (less than 7%) thought their unit would be less close although almost 14% reported that it was. These numbers are reflected in a statistically significant ($p = .003$) difference in mean responses between the Pendleton ($\mu = 3.92$) and Okinawa ($\mu = 3.58$) samples.

<u>Item: "While in Okinawa, the officers in my unit will be."</u> A majority (56%) of marines expected greater strictness from their officers while the unit was in Okinawa and over 54% reported this occurred. However, a significant number neither expected (42%) nor reported (37%) much change in officer strictness. Expectations were very close to reported degree of strictness and there is no real difference in mean responses between the Pendleton and Okinawa samples on this item.

<u>Item: "While in Okinawa, rules and regulations will be enforced."</u> In a manner similar to the previous question a large majority (88%) of respondents expected rules and regulations to be more strictly enforced. Over 82% reported that rules and regulations were followed more strictly while the unit was in Okinawa. Again, expectations were very close to reported

results, and statistical testing suggests there is no real difference between the two samples.

Item: "Compared to here, the weather in Okinawa will be." Marines responded to this question using a five point scale ranging from "much worse" (1) to "much better" (5). Here again, expectations seemed accurate. Over 61% expected worse weather in Okinawa and over 64% reported so. More than 16% of the Pendleton sample had expected better weather, but only 6% of the Okinawa sample perceived the weather as superior to southern California's. There is no statistically significant difference between the Pendleton mean response (2.52) and the Okinawa mean response (2.32).

Item: "Compared to here, the activities and entertainment available for leisure time in Okinawa will be." Using the same answer format as the previous question, over 37% of the Pendleton sample expected leisure time activities to be better, but only 19% of the Okinawa sample reported improvement. Some 33% of the marines had expected activities and entertainment to be worse in Okinawa; a majority (58%) reported that this was so. The mean response from Okinawa (2.43) is lower than the Pendleton mean (3.03) by a statistically significant amount ($p < .001$).

Item: "Compared to here, in Okinawa I will have." This item inquired about friendship, and respondents answered on a scale from "many fewer friends" (1) to "many more friends" (5). Only about 21% of the marines expected to have more friends in Okinawa, however this number almost doubled with over 40% later reporting having more friends. A majority (over 61%) of marines had expected their number of friends to remain the same but less than 41% reported this to be so. Approximately the same percentage of respondents expected and reported having fewer friends. The Pendleton sample mean of 2.98 is statistically different from the Okinawa sample mean of 3.25 at $p = .005$.

Item: "In Okinawa, I will miss my family/relatives." Almost 54% of the marines expected to miss their family and other relatives more while overseas. Some 61% reported that they did miss their families more while in Okinawa. However, a fairly large number expected (40%) and did miss (33%) them about the same. There is no statistically significant difference in mean responses between the two samples.

Item: "Overall, I expect my transfer and adjustment to Okinawa to be." Respondents chose among five answers varying from "very difficult" to "very easy." Over 14% of the Pendleton sample expected a difficult transfer, but only 8% reported difficulty in adjustment. Almost 50% of the marines expected an easy adjustment to Okinawa and a majority (57%) reported it to be so. The difference in mean responses to this question (3.41 versus 3.65), though not large numerically, is not due to chance being statistically significant at $p = .012$.

The differences found between the Okinawa and Pendleton samples in their responses to these 13 items, brings into question the effects that such differences in expectations about Okinawa have upon the adjustment of marines to their transfer. A number of analyses were conducted to examine the relationship between the "reality" of Pendleton expectations (as compared to what it was actually like in Okinawa) and several measures of transfer adjustment. Difference scores (Okinawa minus Pendleton) were used in some of the analyses conducted. However, because of the serious psychometric difficulties associated with the use of difference scores, alternative ANOVA and regression techniques were used which did not depend on difference scores as the primary data base. The result of all of these analyses showed that differences between Pendleton and Okinawa did not add significantly to the prediction of transfer adjustment above that gained by

knowledge of Pendleton expectations or Okinawa evaluations separately. Interestingly, we found (using regression procedures) that certain aspects of adjustment, i.e. intention to re-enlist, reports of heavy drinking, and "getting angry" (measured in Okinawa) were better predicted by Pendleton expectations (R's = .52, .53, and .53 respectively, $p < .05$) than by evaluations of Okinawa measured in Okinawa (R's = .43, .44, and .39 respectively, $p < .05$). An examination of the specific Pendleton items which entered into the regression equations indicated that marines who left Pendleton with certain types of expectations (e.g. the officers will be strict, the entertainment opportunities will be bad, or the Okinawans will be friendly) had greater adjustment difficulties than did marines who left Pendleton with different attitudes about what Okinawa would be like. These Pendleton expectations were related to Okinawa adjustment regardless of what the marines found Okinawa to be like.

SUMMARY AND DISCUSSION

For the 13 questions asked of both Pendleton and Okinawa samples, differences in mean responses to nine were statistically significant ($p < .05$). This number is much higher than would be expected by chance. Differences in responses to another three questions were marginally significant (p between .05 and .10).

For a number of items, expectations concerning Okinawa were higher or more positive prior to transfer than the marines' later assessment of reality in Okinawa. This "negative" change occurred for items dealing with living in Okinawa, the job to be done there, standard of living, unit cohesiveness, and availability of activities and entertainment. On the other hand, some things about Okinawa turned out better than expected on average. For example, marines reported having more friends than they expected.

A number of expectations -- for example, with regard to Okinawans, officer strictness, rules and regulations, weather on Okinawa, and missing family -- were not significantly at odds with later assessment. It would seem that marines had a realistic and accurate view of what to expect in these areas.

Finally, the data indicated that, at least in the case of three adjustment variables, differences between what marines expected Okinawa to be and what Okinawa actually was were not significantly predictive of Okinawa adjustment when compared to data concerning expectations alone. The nature of these findings might indicate a certain self-fulfilling prophecy on the part of some marines (Jones, 1977). Specific expectations about Okinawa lead to difficulty in transfer regardless of the extent to which those expectations are proven accurate by the Okinawa experience.

REFERENCES

Feldman, D.C. A contingency theory of socialization. Administrative Science Quarterly, 1976, 21, 433-452.

Feldman, D.C. The multiple socialization of organization members. Academy of Management Review, 1981, 6, 309-318.

Fisher, C.D., Wilkins, C., & Eulberg, J. Transfer transitions. Report prepared for the Office of Naval Research, Texas A&M University, 1982, TR-ONR-5.

Ilgen, D.W. & Seeley, W. Realistic expectations as an aid in reducing voluntary resignations. Journal of Applied Psychology, 1974, 59, 452-455.

Jones, R.A. Self-Fulfilling Prophecies: Social, Psychological, and Physiological Effects of Expectancies. Hillsdale, NJ: Erlbaum, 1977.

Van Maanen, J. Breaking in: Socialization to work. In R. Dubin (Ed.), Handbook of Work, Organization, and Society. Chicago: Rand McNally, 1976.

Van Maanen, J. & Schein, E.H. Toward a theory of organizational socialization. In B. Staw (Ed.), Research in Organizational Behavior, Vol. I. Greenwich, Conn.: JAI Press, 1979.

Wanous, J.P. Effects of realistic job preview on job acceptance, job attitudes, and job survival. Journal of Applied Psychology, 1973, 58, 327-332.

Zaharia, E.S. & Baumeister, A.A. Job preview effects during the critical initial employment period. Journal of Applied Psychology, 1981, 66, 19-22.

A REDESIGNED PERFORMANCE APPRAISAL SYSTEM FOR NONCOMMISSIONED RANKS IN THE CANADIAN ARMED FORCES

LCDR W. S. Shields, Canadian Forces Personnel
Applied Research Unit, Toronto, Canada[1]

This new system, to be installed in 1983, was based on two years of research by a six-person team. The study used secondary occupational analysis data, content analysis of performance appraisal narratives, psychometric analysis of performance appraisal scores, primary survey data, and longitudinal studies, to identify fourteen performance assessment dimensions, each of which displayed broad applicability across the more than 100 assessee occupations. Unique features of the appraisal system are (a) a new approach to numerical scale anchoring, (b) a built-in score-monitoring system, and (c) the ability to adapt the system to each occupation according to the importance of each assessment factor in that occupation, while maintaining a standard format.

Background

The existing Personnel Evaluation Report (PER) for noncommissioned personnel in the Canadian Forces (CF) has been in use since 1968 and was based on a survey given to 438 Canadian Army personnel in 1966. The survey used Flanagan's (1954) "critical-incident" technique to obtain 23 "summary statements of behaviour", 19 of which were used initially in the PER and 17 of which remain in use. An evaluation of this PER, and its associated orders and instructions, was undertaken in 1980. The need for a revalidation of the PER was dictated both by the number of years since the 1966 study and by the need to derive the content from data not just from soldiers, but also from sailors and airmen.

Research Strategy

A prime consideration in the redesign of the PER was determining the performance dimensions to be assessed. The basic strategy adopted to accomplish this was: (a) identify and eliminate the redundancies, if any, in the existing performance criteria, and (b) identify and fill the gaps in the criteria - performance dimensions that are not now assessed but ought to be.

A. IDENTIFYING REDUNDANCIES

Procedure

Because the most important use of the appraisal form is in making promotion decisions, the ability of the form to predict performance in the next higher rank is essential. Therefore, a study was made of all noncommissioned personnel promoted in 1978, to see which of the 17 performance dimensions (assessed in 1976) were predictive of performance in the next higher rank (assessed in 1980). An equally-weighted score-average was used to estimate an individual's 1980 performance.

Results

All 17 performance dimensions proved to be positively and significantly correlated with performance in the next higher rank, at all levels Corporal (Cpl) through Chief Warrant Officer (CWO). However, this could be partly because the same instrument was used to measure performance in both 1976 and 1980, and also partly because of halo in the appraisal form. In fact, several of the correlations among performance dimensions were fairly high. For example, 13 of them, involving seven dimensions, were greater than 0.71, indicating 50% or more variance overlap between the pairs of dimensions so correlated. The strongest of these was between "knowledge of trade/job" and "ability to apply knowledge", suggesting that supervisors could assess "knowledge" only in its observable form, i.e. its application.

To obtain a reasonably nonredundant subset of the 17 dimensions, forward stepwise regressions to predict performance at the next rank level were performed at each of five career segments, using a program which prohibited entry of variables with negative beta-coefficients. The program was directed to stop as soon as no variable additions, removals, or trades could reduce the standard error of estimate. The results are shown in Table 1.

Note that eight of the 17 dimensions had no unique predictive value at any of the rank levels examined. "Applying job knowledge" was only uniquely predictive at the Corporal (a nonsupervisory) rank. "Initiative" and "Conduct" appeared to have greater importance at more junior levels, "Planning" was important at all supervisory levels, and "Supervision" was uniquely important only at the most senior levels. Communicating (Briefing) was fairly uniformly important through all levels. Its failure to appear at the MWO-CWO level may be an artifact of the relatively small sample there.

[1] The views expressed in this paper are those of the author and not necessarily those of the Department of National Defence. The author acknowledges the highly professional participation in this project of Col M.D. Gates (Project Team Leader), LCol G.M. Rampton PhD, Maj J.P. McMenemy, Capt J.M. McCutcheon, Capt S.T. Halliwell PhD, Lt M.A. Akodie PhD, C.P.O. B.P. Weber, and numerous other members of the National Defence Headquarters and CFPARU staff.

Table 1

RESULTS OF FORWARD STEPWISE REGRESSIONS TO PREDICT PERFORMANCE FOUR YEARS LATER - TWO YEARS INTO THE NEXT HIGHER RANK

Performance Dimension	BETA COEFFICIENTS				
	Cpl-MCpl N=842 R=.29346	MCpl-Sgt N=983 R=.31311	Sgt-WO N=470 R=.37607	WO-MWO N=220 R=.54732	MWO-CWO N=68 R=.42143
1. Preparation & Planning		.06462	.11833	.14178	.19528
2. Delegation					
3. Performance Under Stress		.08236			.12095
4. Cooperation					
5. Command & Self Assertion					
6. Support of Subordinates					
7. Briefing Others	.06680	.07851	.11468	.13275	
8. Knowledge of Trade/Job					
9. Ability to Apply Knowledge	.09671				
10. Adaptability	.06236			.13070	
11. Initiative	.13650	.13300	.11027		
12. Appearance & Bearing				.26118	
13. Supervision				.13899	.30688
14. Ensuring Understanding					
15. Responsibility					
16. Conduct	.06001	.15603			
17. Learning					

B. IDENTIFYING GAPS

Method

A primary source of new PER dimensions was a content analysis of 585 historical PER narratives, approximately 100 in each of the ranks Cpl through CWO. The results obtained were further supported by a psychometric analysis of historical PER scores, analysis of secondary occupational analysis data, and by a survey given to 243 officers and 200 other ranks. The narrative content analysis had several objectives, the primary one being to identify those dimensions of effective performance that pervade all 104 CF military occupations and, hopefully, all noncommissioned rank levels.

Results

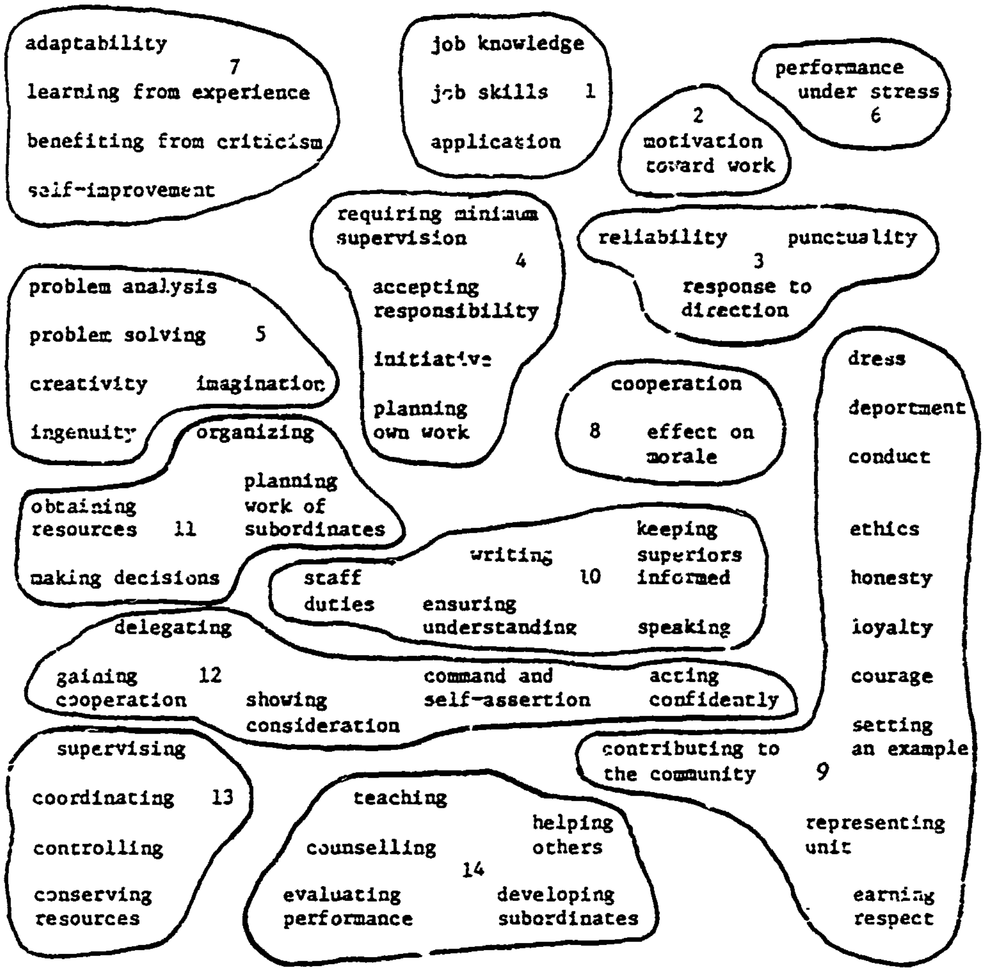
The 585 narratives contained 47 performance dimensions, the present 17 plus 30 others. The analysis tabulated not only the presence of one of these in a given person's narrative, but also quantified the person's assessment in that dimension using a number derived from the adjectives used in the narrative to describe the performance. Thus, it was possible to estimate the correlations among all dimensions except those which were mentioned very rarely. The correlations were expressed as "city block" distances (dissimilarities). A "Johnson" (1967) clustering of these dimensions (using the "maximum" distance criterion), aided by the judgements of a panel of content experts, was used to group the dimensions into the 14 clusters shown in Appendix 1. The appendix also contains the frequency of mention of the dimensions in the narratives, and other data indicative of their relative importance. Figure 1 diagrams a crude two-dimensional scaling of the dimensions, and serves to emphasize that other cluster solutions, some perhaps superior to that chosen, are possible. However, the solution chosen seemed to produce relatively few classification ambiguities; therefore the clusters of Figure 1 were chosen as the 14 assessment factors in the new PER.

A factor analysis of historical PER scores had earlier produced only three prime factors that were consistent across all occupations: (1) Does own work, (2) Influences others, and (3) Supervises others. These factors were used to place the 14 clusters into three groups, as shown in Appendix 1. Group 2, "Influences others" has the unique property that all its assessment factors can, like those in Group 1, be rated for a person who has no subordinates, and yet the factors constitute qualities of prime importance to leadership. Thus, they should constitute valuable indicators of leadership potential.

A philosophy of cluster naming and assignment that aided greatly in reducing the number of clusters (assessment factors) to 14 was to express, wherever possible, the cluster name as an <u>outcome</u>. For example, a single dimension "earning respect" was used in Cluster 9 to aggregate several specific qualities. Without this kind of consolidation, the PER would have had far too many assessment factors to serve the practical needs of the CF.

Figure 1

PER NARRATIVE CONTENT DIMENSIONS



Usually, the selection of assessment factors is a sampling process whereby a few of the more important dimensions of performance are rated and taken as representative of the much larger number of relevant and assessable ones. It is anticipated that the use of a clustering rather than a sampling philosophy will result in a more comprehensive set of appraisal criteria.

Although the content analysis revealed some significant changes in content with increasing rank, these were not large enough to constitute a persuasive argument for the use of separate forms for junior and senior ranks, especially in view of the capacity of the new form, like that of the old one, to permit omission of supervisory factors for personnel who have no subordinates.

C. PER DESIGN FEATURES

The new PER form is shown in Appendix 2. It consists of a single 8 1/2" by 12" machine-readable sheet. The instructions which accompany the form contain a table of "word-pictures" that describe each level of performance for each factor. Levels 6 and 7 are combined in these descriptions; therefore there are 84 word-pictures in all. As an example, the word-pictures for Performance Factor 12 "Gaining Cooperation" are shown in Table 3. The word-pictures were composed by a panel of experts, rather than using conventional BARS (Behaviourally Anchored Rating Scale) techniques; however, their clarity and ordinal properties are being verified through field trials.

Table 3

EXAMPLE OF A ROW IN THE WORD-PICTURE TABLE

12. GAINING COOPERATION
Consider: showing considera[illegible] or subordinates; giving praise and [illegible]ticism; effect on subordinates' motivati[illegible]

4	5
Makes unreasonable demands; seldom praises and often criticizes unfairly; demoralizes subordinates.	Has limited success in motivating others; sometimes unmindful of subordinates' well-being.

6/7	8
Obtains willing cooperation through fairness, reasonable expectations, and giving credit when due.	Subordinates follow and respond eagerly.

9	10
Praises and criticizes wisely; impressive effect on subordinates' motivation.	Even in the most difficult circumstances, inspires the loyalty of all; subordinates strive to earn his/her praise.

The 10-point scale was chosen because of its common use in Canadian society, particularly in evaluating student work in education and training. The scale also aligns very well with the concept of efficiency. It is truncated because pre-selection of CF applicants, together with further weeding out during training, results in the virtual absence of personnel with less than 40% efficiency.

A unique feature of the form is its built-in monitoring system. Unit COs will be provided with CF norms regarding the percent usage of the highest two rating categories. If Unit usage of these categories departs from these norms within acceptable limits, PERs will be monitored at headquarters only for technical errors. If, however, deviations are unacceptable and not accompanied by documentary evidence of either preferred manning or extraordinary unit effectiveness, the PERs will be subjected to a very stringent monitoring in the case of small units, and returned to unit (RTU) en masse in the case of larger units. COs who neither substantiate scores adequately nor bring them into acceptable conformity after a second RTU will be liable to have all PERs from their units stamped in red with the correction factor that merit boards will be advised to apply to each person's aggregated score.

A feature of the PER system that has been deferred, pending the gathering of additional data, is a plan for differential weighting of the criteria according to their importance in the particular rank/trade combination being assessed. The aggregated performance score, if this feature is implemented, will then be the scalar product of the score vector with the importance vector. It is envisioned that a "potential" score will also be calculated, which will be the scalar product of the score vector with the importance vector in the next higher rank in the trade assessed. The importance data will be gathered after the new PER has been in place long enough for personnel to have a good familiarity with the 14 newly defined performance factors.

Summary

The assessment factors in the new PER retain five of the previous 17 virtually unchanged. Two other factors constitute relatively minor modifications of previous ones. The remaining seven are either completely new or substantially different from those in the old form. The revised set of factors has a solid research foundation, can be claimed to represent all CF environments, and reflects the current value systems of CF supervisors, reviewing officers and commanding officers.

References

Flanagan, J.C. The critical incident technique. Psychological Bulletin, Vol 51, No. 4 (1954), 327-358.

Johnson, S.C. Hierarchical Clustering Schemes. Psychometrika , Vol 32 (September, 1967), 241-254.

APPENDIX 1

CLUSTERING OF NARRATIVE CONTENT DIMENSIONS

Cluster Names

Does Own Work
1. Applying job knowledge and skills
2. Motivation toward work
3. Response to direction
4. Initiative
5. Problem solving
6. Performance under stress
7. Learning

Influences Others
8. Effect on group morale
9. Earning respect
10. Communicating

Supervises Others
11. Planning work of subordinates
12. Gaining cooperation
13. Coordinating/controlling
14. Developing subordinates

PERFORMANCE DIMENSIONS	FREQUENCY OF MENTION IN THE 585 NARRATIVES	FURTHER STRONG SUPPORT (S=Survey OA=Occupational Anal)	CLUSTER MEMBERSHIP
EXISTING			
1. Preparation and Planning	328	S, OA	4, 11
2. Delegation	135		12
3. Performance under Stress/Pressure	208	S, OA	6
4. Cooperation	392	S, OA	3, 8
5. Command and Self Assertion	129	S	10, 12
6. Support of Subordinates	228	S	14
7. Briefing Others	236	S, OA	10
8. Knowledge of Trade/Job	450	S	1
9. Ability to Apply Knowedge	536	S	1
10. Adaptability	272	S, OA	7
11. Initiative	261	S, OA	4
12. Appearance and Bearing	296		3, 9
13. Supervision	250	S	13
14. Ensuring Understanding of Assignments	78	[illegible]	10, 12
15. Responsibility	307	S	3, 4
16. Conduct	210		3, 9
17. Learning from Experience	128	S	7
NEW			
18. Communicated efficiently in writing	75	OA	10
19. Acted confidently	101		10, 12
20. Gained the cooperation of others	221	OA	12
21. Acted courageously	6		9
22. Accepted criticism	9		7
23. Made sound decisions	50	S, OA	5, 11
24. Observed military ethics	7		9
25. Acted honestly	13	S	9
26. Displayed creativity, ingenuity and imagination	100		4, 5
27. Displayed loyalty to the unit and to the Forces	97		9
28. Was beneficial to group morale	119	OA	8
29. Was well motivated towards own work	384	S	2
30. Analysed and solved problems	151	S, OA	5
31. Was punctual	12		3
32. Was reliable	130	S, OA	3, 4
33. Obtained needed resources	9		11
34. Used resources efficiently	44		13
35. Responded to direction	54	OA	3
36. Pursued self-improvement	122		7
37. Required a minimum of supervision	146	OA	3, 4
38. Earned the respect of others	180		9
39. Performed staff duties	24		5, 10
40. Contributed to the community	98		9
41. Kept superiors informed	37		10
42. Organized work	141	OA	4, 11
43. Represented unit	19	OA	9
44. Coordinated activities	41		13
45. Others sought his advice	38		9, 14
46. Was responsible	69		3, 4
47. Set a good example	1		8, 9

APPENDIX 2

National Defence — Défense nationale

CONFIDENTIAL (When any part other than section 2 completed)

PERFORMANCE EVALUATION REPORT
OTHER RANKS

CONFIDENTIEL (lorsque toute partie autre que la section 2 est remplie)

RAPPORT D'APPLICATION DU RENDEMENT
NON OFFICIERS

SECTION 1
PERSONAL INFORMATION
RENSEIGNEMENTS PERSONNELS

1A. SURNAME-NOM	INITIALS-INITIALES	SIN-NAS	MOC-CEM	NAME OF UNIT-NOM DE L'UNITE

1B. MARITAL STATUS-ETAT CIVIL

☐ SINGLE / CELIBATAIRE ☐ MARRIED / MARIE E ☐ SEPARATED / SEPARE E ☐ DIVORCED / DIVORCE E ☐ WIDOWED / VEUF VE

1C. LOCATION OF DF & E-EMPLACEMENT DES PM ET E

(Date moved - Date de demenagement)

1D. DEPENDANT CHILDREN-DETAILS CONCERNANT LES ENFANTS A CHARGE

	1	2	3	4	5	6	7	8	9	10
SEX (M OR F) / SEXE (M OU F)										
AGE / AGE										
SCHOOL GRADE / ANNEE SCOLAIRE										
LANGUAGE OF INSTRUCTION (E/F) / LANGUE D'INSTRUCTION (A/F)										

1E. IF SPOUSE MILITARY, GIVE DETAILS-SI LE CONJOINT EST MILITAIRE, DONNER LES DETAILS

SURNAME-NOM — INITIALS-INITIALES

SIN-NAS — NAME & LOCATION OF UNIT / NOM ET LIEU DE L'UNITE

MOC-CEM

1F. COURSES COMPLETED SINCE LAST REPORT (IN AND OUT OF SERVICE)
COURS TERMINES DEPUIS LE DERNIER RAPPORT (MILITAIRES ET NON MILITAIRES)

1G. COURSES DESIRED-COURS DESIRES

(1) SECOND LANGUAGE TRAINING-INSTRUCTIONS EN LANGUE SECONDE

DESIRED-DESIREE ☐ NOT DESIRED-NON DESIREE ☐

(2) OTHER COURSES-AUTRES COURS

1H. FACTORS AFFECTING FUTURE POSTINGS
FACTEURS INFLUANT SUR LES AFFECTATIONS FUTURES

☐ DEPENDANTS EDUCATION / ETUDES DES PERSONNES A CHARGE
☐ DEPENDANTS MEDICAL / PROBLEME MEDICAL DES PERSONNES A CHARGE
☐ HOME OWNER / PROPRIETAIRE DE MAISON
☐ SPOUSE EMPLOYMENT / EMPLOI DU CONJOINT

Location-Endroit ____________

OTHER-AUTRE

1J. POSTING PREFERENCES-AFFECTATIONS PREFEREES

(1)

(2)

(3)

1K. NARRATIVE LANGUAGE REQUESTED
LANGUE DEMANDEE POUR LA NARRATION

☐ ENGLISH-ANGLAIS
☐ FRENCH-FRANCAIS
☐ EITHER-L UN OU L AUTRE

1L. SECTION 1. CERTIFIED CORRECT-SECTION 1. CERTIFIEE CORRECTE

(Signature) (Date)

SECTION 2 — **2A. IDENTIFICATION**

2C TYPE OF REPORT / GENRE DE RAPPORT: SPECIAL / SPECIAL; UN COURSE / SUR COURS; POSTING / AFFECTATION; EXTENDED ANNUAL / ANNUEL PROLONGE; ANNUAL / ANNUEL

2D SUPERVISOR / CONVEILLANT: CIVILIAN / CIVIL; RANK / GRADE: BGEN, COL, LCOL, MAJ, CAPT, LT, 2LT, CWO, MWO, WO, SGT, MCPL; MOC CEM

2E REPORTING PERIOD / PERIODE D'EVALUATION: FROM / DE — MO / MOIS, YEAR / ANNEE; TO / A — MO / MOIS, YEAR / ANNEE

2F UIC OF REPORTING UNIT / CNU DE L'UNITE EVALUATRICE

RANK GRADE: CWO / ADJUC; MWO / ADJUM; WO / ADJ; SGT / SGT; MCPL / CPLC; CPL / CPL; PTE / SDT

SOCIAL INSURANCE NO / N° D'ASSURANCE SOCIALE

MOC CEM

SURNAME NOM

2G POSTING PREFERENCES-AFFECTATIONS PREFEREES

1

2

3

CONFIDENTIAL (When any part other than section 2 completed) - CONFIDENTIEL (Lorsque toute partie autre que la section 2 est remplie)

SECTION 3

3A. QUANTITATIVE ASSESSMENT ÉVALUATION QUANTITATIVE

NOT OBSERVED / NON OBSERVÉ(E)

UNSATISFACTORY-INSATISFAISANT(E) · MARGINAL-PASSABLE · SATISFACTORY-SATISFAISANT(E) · GOOD-BON(NE) · SUPERIOR-SUPÉRIEUR(E) · OUTSTANDING-HORS PAIR

DOES OWN WORK			FAIT SON PROPRE TRAVAIL	
1 APPLYING JOB KNOWLEDGE / SKILLS		①②③④⑤⑥⑦⑧⑨⑩	APPLICATION DES CONNAISSANCES APTITUDES AU TRAVAIL	1
2. MOTIVATION TOWARD WORK		①②③④⑤⑥⑦⑧⑨⑩	MOTIVATION AU TRAVAIL	2
3 RESPONSE TO DIRECTION		①②③④⑤⑥⑦⑧⑨⑩	RÉACTION AUX DIRECTIVES	3
4 INITIATIVE		①②③④⑤⑥⑦⑧⑨⑩	INITIATIVE	4
5 PROBLEM SOLVING	Ⓝ	①②③④⑤⑥⑦⑧⑨⑩	RÉSOLUTION DES PROBLÈMES	5
6 PERFORMANCE UNDER STRESS	Ⓝ	①②③④⑤⑥⑦⑧⑨⑩	RENDEMENT SOUS TENSION	6
7 LEARNING	Ⓝ	①②③④⑤⑥⑦⑧⑨⑩	APPRENTISSAGE	7
LEADERSHIP			COMMANDEMENT	
8 EFFECT ON GROUP MORALE		①②③④⑤⑥⑦⑧⑨⑩	INFLUENCE SUR LE MORAL DU GROUPE	8
9 EARNING RESPECT		①②③④⑤⑥⑦⑧⑨⑩	GAIN DE RESPECT	9
10. COMMUNICATING		①②③④⑤⑥⑦⑧⑨⑩	APTITUDE À COMMUNIQUER	10
11 PLANNING WORK OF SUBORDINATES	Ⓝ	①②③④⑤⑥⑦⑧⑨⑩	PLANIFICATION DU TRAVAIL DES SUBALTERNES	11
12. GAINING COOPERATION	Ⓝ	①②③④⑤⑥⑦⑧⑨⑩	CAPACITÉ D'OBTENIR LA COLLABORATION	12
13 COORDINATING / CONTROLLING	Ⓝ	①②③④⑤⑥⑦⑧⑨⑩	COORDINATION / CONTRÔLE	13
14 DEVELOPING SUBORDINATES	Ⓝ	①②③④⑤⑥⑦⑧⑨⑩	PERFECTIONNEMENT DES SUBALTERNES	14

SECTION 7 PER MONITORING-CONTRÔLE DES RAP

PASSED MONITORING - A PASSÉ LE CONTRÔLE ____ ○

FIRST RTU / PREMIER RAU ____ ○

SECOND RTU / DEUXIÈME RAU ____ ○

ACCEPTED WITH-ACCEPTÉ AVEC-OBSERVATIONS ____ ○

SCORE CORRECTION BASED ON CF SCORING NORMS / CORRECTION DES COTES BASÉE SUR LES NORMES DES FC

ADD ☐ AJOUTER

SUBTRACT ☐ SOUSTRAIRE ____ POINTS

3B. PHYSICAL FITNESS TEST / TEST DE CONDITION PHYSIQUE — PASSED-PASSÉ ○ FAILED-ÉCHOUÉ ○ EXCUSED-EXCUSÉ(E) ○

3C. NUMBER OF SUBORDINATES OF ASSESSEE / NOMBRE DE SUBALTERNES DE L'ÉVALUÉ(E)

AS SUPERVISOR / COMME SURVEILLANT ☐ AS REVIEWING OFFICER / COMME RÉVISEUR ☐

3D. 1st PER-1er RAP — BY THIS UNIT-PAR CETTE UNITÉ ○ BY THIS SUPERVISOR-PAR CE(TTE) SURVEILLANT(E) ○

3E. SCORE AVERAGE / MOYENNE DE POINTAGE ▶ ◀

SECTION 4 4A. NARRATIVE ASSESSMENT-ÉVALUATION NARRATIVE

DURING THE LAST / AU COURS DES DERNIERS ☐ MONTHS THE MEMBER HAS BEEN EMPLOYED AS / MOIS A ÉTÉ EMPLOYÉ(E) EN QUALITÉ DE ▶

PERFORMANCE OF DUTIES
ACCOMPLISSEMENT DES TÂCHES

4B. SUPERVISOR SURVEILLANT(E) (CF48CA) — NAME RANK APPOINTMENT & SIN-NOM GRADE POSTE ET NAS

☐ COMPLETE AND CORRECT / COMPLET ET CORRECT

☐ UNDERGOING AMENDMENT / SUBIT MODIFICATION(S)

SIGNATURE | DATE

4C. REVIEWING OFFICER OFFICIER RÉVISEUR — NAME RANK & APPOINTMENT NOM GRADE ET POSTE

SIGNATURE | DATE

SECTION 5 COMMANDING OFFICER-OFFICIER COMMANDANT

5A. RECOMMENDATIONS-RECOMMANDATIONS

(1) PROMOTION

○ NOT SUITABLE NON CONVENABLE
○ LIKELY TO BECOME SUITABLE POURRAIT DEVENIR CONVENABLE
○ SUITABLE CONVENABLE
○ HIGHLY SUITABLE HAUTEMENT CONVENABLE

(2) REENGAGEMENT - RÉENGAGEMENT

○ RECOMMENDED RECOMMANDÉ
○ NOT RECOMMENDED NON RECOMMANDÉ
○ NOT APPLICABLE (PS) NE S'APPLIQUE PAS (PS)

5B. REMARKS-REMARQUES

NAME & RANK NOM ET GRADE | SIGNATURE | DATE

SECTION 6 ASSESSEE-ÉVALUÉ(E) — SECTIONS 4 AND 5 HAVE BEEN READ AND DISCUSSED / SECTION 4 ET 5 ONT ÉTÉ LUS ET DISCUTÉ(E)S — SIGNATURE ▶ DATE ▶

CONFIDENTIAL (When any part other than section 2 completed) - CONFIDENTIEL (Lorsque toute partie autre que la section 2 est remplie)

AD P000908

Revision of the Organizational Assessment Package System: Improvements in Assessing Air Force Organizations

Major Lawrence O. Short, Ph.D.
Chief, Research Operations
Research and Analysis Directorate
Leadership and Management Development Center
Maxwell AFB AL 36112

The Organizational Assessment Package (OAP) and its accompanying consulting process have been in operation since 1979. Until recently, use of the OAP was predicated on developmental work reported in Hendrix and Halverson (1979a; 1979b) and Hendrix 1979. Lessons and experiences from more than two years of field use needed to be incorporated into the instrument and its associated software systems. This became the purpose of a research project to both look closely at the existing OAP and to point the way toward the needed revisions in many OAP system elements supporting the consulting process.

This paper is organized around two major issues. First is the description of the current OAP, the associated consulting process, and some research results pertaining to the OAP. Second is a description of the major elements in the revision and how the changes will improve parts of the OAP system.

The Current System

Description of the Instrument and Consulting Process

The OAP is a 109 question survey designed jointly by the Air Force Human Resources Laboratory and the Leadership and Management Development Center (LMDC) to aid the LMDC in its mission to: (a) provide management consulting services to Air Force commanders upon request, (b) to provide leadership and management training, and (c) to conduct research on Air Force systemic issues with information within the accumulated data base.

Administration of the survey is the first step in the consultation process. The survey is given to a stratified random sample of the organization to which LMDC has been invited. The results of the survey are an important feature in the assessment of the organization. The results are handled in a confidential manner between LMDC and the client. After approximately five to six weeks for analysis, consultants return to the organization to provide feedback of data to commanders and supervisors.

When organizational problems are encountered, a consultant and supervisor develop a management action plan designed to resolve the problem at that level of the organization. Within six to nine months, the consulting team returns to readminister the survey instrument as a means to help assess the impact of the consulting process.

The data from each OAP administration effort are stored in a cumulative data base currently containing over 100,000 records for research purposes. These data are aggregated by work group codes developed for this instrument. The data may be recalled by demographics such as personnel category, age, sex, Air Force Specialty Code (AFSC), pay grade, time in service, and educational level. Through factor analysis, the 93 attitudinal items are combined into

factors which cover job content, job interferences, and various types of supervisory and organizational areas. OAP factor names are presented in Figure 1.

Skill Variety	Advancement/Recognition
Task Identity	Management-Supervision
Task Significance	Supervisory Communications Climate
Job Feedback	Organizational Communications Climate
Work Support	Perceived Productivity (Work Group Effectiveness)
Need for Enrichment (Job Desires)	Job Satisfaction
Job Performance Goals	Job Related Training
Pride	General Organizational Climate
Task Characteristics	
Task Autonomy	
Work Repetition	
Desired Repetitive Easy Tasks	

Figure 1. OAP Factor Names

Some Selected Research

The current state of the OAP as well as need for selected revisions can be seen by reviewing some OAP related research. First, Short and Hamilton (1981) provided evidence of the factor-by-factor reliability of the instrument. Prior to this study, OAP factors were expected to be internally consistent as assessed by a Cronbach's alpha procedure and were expected to retain significant test-retest correlations across both five week and six month time intervals. It was further expected that the six month correlations would be lower than those for the five week interval because of both the longer interval and the necessity that factors be sensitive to actual organizational changes rather than being artificially rigid. These expectations were confirmed with the exception of some of the two or three item factors. Therefore, reliability for the primary OAP factors was shown to be acceptable to excellent.

In addition, Short and Wilkerson (1981) offered support for the group differences aspect of OAP construct validity. Hypotheses tested were stated at three levels. First, it was anticipated that all OAP factors would be sensitive enough that between group variance would exceed within group or error variance across functional area groupings. Corresponding null hypotheses of no differences among functional areas were stated for every factor. Second, based in part on the work of Conlon (1980) and in part on consultants' observations of task, climate, productivity and leadership patterns Air Force wide, it was expected that factors dealing with perceptions of task would show the widest variation across functional areas. Similarly, it was expected that perceptions dealing with leadership function and style would be most consistent and show the least variation across functional area groupings. Since perceptions of climate as defined in the OAP may be related to perceptions of task, it was expected that climate factors would show variations second only to task. Finally, it was expected that perceived productivity, dependent to a degree on all the other three, would show more variation than leadership factors but less variation than task or climate factors. These logical factor groupings and the hypothesized direction of differences were summarized as follows:

Perceptions of Task	>	Perceptions of Climate	>	Perceptions of Productivity	>	Perceptions of Leadership

Finally, specific pairwise differences between groups and direction of differences by factor across functional area groupings were hypothesized where information was available upon which to base such hypotheses.

Results showed differences by OAP factor across major functional area groupings were consistent and strong. These differences also held across logical groupings of factors. Results were more equivocal, however, concerning specific pair comparisons within factors. These results provide strong support for one aspect of OAP construct validity, but they also showed a need for further revision of the instrument, especially in regard to the two or three item factors.

Next, Webster (1982) approached the construct validity of OAP leadership and organizational climate indices by use of the multitrait - multimethod approach. This method allows assessment of both convergent (Are OAP results related to outside criteria where relationships would be expected?) and discriminant validity (Do low relationships exist where those would be expected?). The criterion measures in the study were the leadership and climate measures from the Survey of Organization (SOO) (Taylor and Bowers, 1972). The SOO is considered a classic in organizational assessment and is described by Nadler (1977) as, "an example of a comprehensive and thoroughly developed instrument, . ." (p. 128). Webster noted that the results clearly indicate significant convergent validity for the OAP, while discriminant validity is also present but is less consistent. He also noted the high intercorrelations between the leadership and climate factors of both instruments as an indicator of possible methods variance which could be addressed by some instrument revision. Thus, while results were again largely positive and encouraging, some revision and sharpening were indicated.

Finally, Hightower and Short (1982a; 1982b) studied consistency of the OAP factors across selected functional area and demographic groups. In order to study factor consistency, responses to the pre-intervention OAP were drawn from the data base and aggregated by major functional area and demographic groupings. The functional area groupings were wing/group staff, resources, maintenance, operations, medical, missiles, communications and unique, a category containing people in organizations with scientific and technical orientations. The demographic groupings included sex, personnel category (officer, enlisted, civilian), and race (white, black, hispanic and other). In addition, factor structure from Survey Time 1 pre-intervention results was compared to Survey Time 2 post-intervention results.

Analyses were accomplished using a principal components solution with a varimax rotation. While principle components analysis and factor analysis are not the same, for sake of simplicity, the term "factor" will be used to refer to resulting statistical groups of items in the remainder of the paper. Each variable was assigned to a factor (component) based on two criteria: loadings above .3 and highest loading. With the exception of the male-female and Survey Time 1 - Survey Time 2 comparisons which were done directly, all other comparison were accomplished by comparing the factor solution for a specific group to the factor solution for the OAP data base exclusive of that group. Three procedures were used to make all comparisons: the congruence coefficient (CC),

the salient variable similarity index (S) and root mean square (RMS). Logical comparison of the results of all three procedures was thought to provide a more precise estimate of the extent of factor matches than could be obtained from any single procedure.

The revised OAP factor solution was slightly different than that currently in use. The revised solution, consisting of thirteen interpretable factors, proved to be consistent across demographic and functional area groups regardless of analysis procedure used. Further, when variables were included only in the factor on which they loaded most highly, examination of factors across the groups studied revealed that the factors consistently contained the same variables. This was especially true when only the variables which loaded strongly on a factor (> .40) were considered.

In general, then, results provided support for the consistency of the OAP revised factor structure across both functional area and demographic groups. Further, this consistency was observed regardless of the method of computing factor matching values. The high values shown when comparing Survey Time 1 with Survey Time 2 results were especially encouraging, since they indicate a high degree of instrument stability across a six month consulting intervention. This finding is especially important when combined with group difference studies. Taken together, these results show an excellent combination of stability, consistency, and sensitivity to change that supports the use of the OAP as both a data gathering and evaluation instrument and point the way for revising and refining the OAP factor structure.

The Revised System

In its simplest form, the revision consists of three major elements: OAP item and factor content, scan sheet and feedback package redesign, and an expanded work group coding system. In regard to the instrument, several additional demographic items will be added. These include items on professional military education, TDY requirements, family information, pay, source of commissioning, technical school training and a revised career intent item. Attitudinal items will be expanded slightly overall and will be summarized by 14 factors (technically components since the "factors" were derived by a principal components analysis). The supervision and organizational climate factors did not separate and will be combined. In addition, new factors measuring job related stress and intergroup conflict will be added and the training factor greatly expanded. Finally, the pride and job satisfaction factors did not separate and will be combined into a job role pride and satisfaction factor. The revised factor structure is contained in Figure 2.

Job Performance Goals	Job Related Stress
Task Characteristics	Supervision
Task Autonomy	Advancement
Work Repetition	Intergroup Conflict
Job Related Training	Work Group Effectiveness
Work Support	Job Role Pride and Satisfaction
Work Interferences	Organizational Climate

Figure 2. Revised OAP Factor Structure Names

The scan sheet and feedback package will be revised consistent with the instrument revision. The scan sheet will have spaces for expanded demographic responses and space for matching code elements which now must be placed in item response positions. These codes are crucial, since they provide a way of linking OAP responses to responses on an additional survey without identifying the respondent. Scan sheets will also be color coded by type of survey to reduce possibility of coding errors. The feedback package is a computer generated document provided to each supervisor who has four or more people from his/her work group respond to the OAP with valid information. Currently, the package provides means, standard deviations and frequency distributions by OAP items and factors. The revised package will include several new elements including an expanded presentation of OAP attitudinal and demographic items and the possibility of a computer graphics generated display of OAP items and factors on which a work group scored lowest. This will allow both consultant and supervisor to more accurately and quickly diagnose work group problems and to propose appropriate interventions and action plans.

Finally, the work group coding system will be expanded beyond its present format. A work group code is a unique combination of alphabetic characters and numeric digits that identifies a functional element within an organization. The code also allows direct comparison of a group with like groups in the data base from other Air Force units. The new coding system will allow more precise coding of a work group and allow groups to be specifically coded down to the lowest level of the organization. This change will greatly help the accuracy and precision of the data base in identifying and comparing specific groups for consulting or Air Force systemic analysis purposes.

A Final Comment

The "bottom line" purpose of the revision was the improvement of a system that was already working well. The elements that have been included should do exactly that. More precisely measuring attitudinal and demographic factors, expanding the way results are returned to supervisors, and more precisely coding all work groups down to the lowest organizational level should be immense help to LMDC's management consultation services in our goal of making a good Air Force better.

References

Conlon, E. J. Investigations of behavioral consultation in the Air Force. In USAF Summer Faculty Research Program Research Reports, Vol 1. Washington, D.C.: The SCEEE Press, 1980. Pp. 14-1 to 14-75.

Hendrix, W. H. Organizational assessment indices of effectiveness (AFHRL-TR-79-46). Brooks AFB, TX: Air Force Human Resources Laboratory, 1979.

Hendrix, W. H., & Halverson, V. B. Organizational survey assessment package for Air Force organizations (AFHRL-TR-78-93). Brooks AFB, TX: Air Force Human Resources Laboratory, 1979. (a)

Hendrix, W. H., & Halverson, V. B. Situational factor identification in Air Force organizations (AFHRL-TR-79-10). Brooks AFB, TX: Air Force Human Resources Laboratory, 1979. (b)

Hightower, J. M., & Short, L. O. Stability of the Organizational Assessment Package factorial validity across groups. Paper presented at the 90th Annual Convention of the American Psychological Association, Washington, D.C., 1982. (a)

Hightower, J. M., & Short, L. O. Factor stability of the Organizational Assessment Package (LMDC-TR-82-1). Maxwell AFB, AL: Leadership and Management Development Center, 1982. (b)

Nadler, D. A. Feedback and organization development. Reading, MA: Addison-Wesley, 1977.

Short, L. O., & Hamilton, K. L. An examination of the reliability of the Organizational Assessment Package (LMDC-TR-81-2). Maxwell AFB, AL: Leadership and Management Development Center, 1981.

Short, L. O., & Wilkerson, D. A. An examination of the group differences aspect of the construct validity of the Organizational Assessment Package. Paper presented at the 23rd Annual Conference of the Military Testing Association, Washington, D.C., October, 1981.

Taylor, J. C., & Bowers, D. G. Survey of organizations. Ann Arbor: Institute for Social Research. University of Michigan, 1972.

Webster, A. H. An assessment of the construct validity of the Leadership and Management Development Center's Organizational Assessment Package (Report Number 82-2640). Maxwell AFB, AL: Air Command and Staff College, 1972.

RETRAINED AIRMEN: VOLUNTEERS VERSUS NON-VOLUNTEERS

Mary J. Skinner
Air Force Human Resources Laboratory
Manpower and Personnel Division, Brooks AFB, TX

Selective (non-voluntary) retraining was identified as a special interest issue in a Request for Personnel Research submitted by managers of the Airman Retraining Program to the Air Force Human Resources Laboratory. The Airman Retraining Program processes 10,000 to 12,000 actions per year which retrain enlisted personnel from one Air Force Specialty (AFS) to another. The majority of the actions are initiated voluntarily by airmen interested in retraining into a second specialty. Other actions, termed selective or non-voluntary retraining, are taken without the concurrence of the individual. Selective retraining is used primarily to fill shortages or reduce manpower overages in specific AFSs. Airmen who disqualify or are unsuited for duty in their awarded specialties are also managed by selective retraining. Most airmen identified as candidates for selective retraining are given notification of the pending action and the opportunity to request voluntary retraining to another AFS. Those few who elect not to exercise the voluntary option are then subject to selective retraining (AFR 39-4, 1979). Historically, Air Force records indicate that selective retraining accounted for less than one percent of all retraining actions processed annually between FY79-82. Because of the possible negative impacts of non-voluntary retraining and anticipated increases in the numbers of non-volunteers, selective retraining issues were addressed as part of a multiphased research effort designed to evaluate the progress of retrainees in their new military specialty. (For an overview of the entire research project, readers are referred to an earlier publication (Skinner, 1981). The objective of the current study was to evaluate the impact of retraining under voluntary or selective conditions on job attitudes, work assignments, performance and adjustment in the second occupation. Policy concerns specific to selective retraining were addressed, as well.

Information for the study was obtained from retrainees and their supervisors who responded to inquiries about selective retraining issues during a field survey in 1980. The survey approach was also used to overcome a major obstacle to the study objective: identification of the selective retrainees. Historical records were thought to be incomplete in distinguishing between selective and voluntary actions. The field survey provided an alternative data source and a means of capturing the individual retrainee's perception of whether he/she considered the job change to have been voluntary or involuntary.

Method

Subjects

Subjects were enlisted personnel who had retrained between July 1973 and August 1979. A stratified-random sample of 20,968 retrainees was selected (see Skinner, 1981 for sampling strategy details). Due to the

relatively small number of selectees identified on historical retraining records, these cases were deliberately oversampled. Later, administrative constraints limited the mail-out to a final sample of 18,065 retrainees and to their first-line supervisors.

Questionnaires

Survey topics and items were identified from literature on the retraining system and from discussions with management personnel. Two questionnaires, one for retrainees and one for supervisors, were developed. The instruments contained standardized items with multiple-response, forced choice alternatives. Most response options were presented in rating scale form.

Questionnaire items and topics pertinent to the study of selective retraining included personal and demographic information and reasons for retraining. Retraining effects on job attitudes and work assignments were evaluated using measures of job satisfaction, perceived use of talents and training, and opportunity for assignment to responsible positions. Performance and ability were assessed by items on quality of work, job knowledge, and supervisory skills. Information on attitudes, motivation, and interpersonal relations was collected to reflect adjustment to the new occupation. The supervisors' questionnaire was designed to collect appraisals of performance, ability, and adjustment. For comparison purposes, supervisors were asked to rate both retrainees and non-retrainees on the factors. Policy concerns were addressed by soliciting opinions about the impact of selective retraining under different conditions.

Analysis

Descriptive statistics for the items (frequency, percentage, mean, standard deviation) were obtained on cases with valid data entries. Tests of significance were conducted with Student t statistics using the Bonferroni technique to control Type I error (α) per family of comparisons (Miller, 1966).

Results

A final analysis sample of 12,827 retrainees remained after data editing. About 79% (N = 10,122) of the retrainees reported that they considered their retraining to have been voluntary. The remaining 21% (N = 2,705) described themselves as involuntary retrainees. Because of possible inaccuracies in historical personnel files, the volunteer/non-volunteer classification reported by the retrainees was subsequently used to categorize data from supervisors into separate analysis groups. Ratings from supervisors of 9,263 voluntary and 2,457 non-voluntary retrainees, as well as of 5,237 non-retrained airmen were analyzed.

The retrainees' own reports of their status were compared with the voluntary or selective identity code on historical personnel records. Table 1 shows the percentage of agreement/disagreement between the two

data sources for the total sample of retrained respondents. Self-reports corresponded with personnel records for 80% of the cases. 77% agreement was found for the volunteer code and 3% for the selectee code. The most notable feature of the discrepancies was that 90% of those who disagreed were retrainees who perceived their retraining to have been selective, but whose personnel records identified them as volunteers (i. e., the 18% of the 20% of cases whose perceptions disagreed with historical files).

Background Information/Retraining Circumstances

Volunteer and selectee groups were similar with respect to occupational and demographic characteristics. The present specialty change was the first retraining experience for the majority of the selectees (74%) and the volunteers (77%). On the average, they had been assigned to the earlier specialty about the same amount of time before retraining. Over 70% of the volunteers and of the selectees had four or fewer years of experience in their current, retraining specialty. Demographic data indicated that the two groups were racially mixed in equivalent proportions and included both male and female enlistees. The majority had completed at least a high school education. Military grades were also comparable; 85% of both groups were in grades E4 through E6.

Views of the circumstances surrounding the retraining experience and reasons for retraining distinguished the selectees and volunteers. Nine of 10 selectees reported that their retraining actions were initiated by the Air Force, and most felt that the job change was effected primarily to satisfy Air Force needs (68%). Selectees, more often than volunteers, reported that their retraining was due to disqualification for the earlier AFS for medical reasons, loss of security clearance, or poor performance. A higher percentage of the selectees retrained, because personnel overages, equipment phase-outs, or CONUS/oversea manpower imbalances were experienced in their former AFSs Volunteers provided divergent descriptions of retraining reasons and circumstances. Over 80% of the volunteers reported that they had initiated their retraining actions. They felt that they were retrained primarily to fulfill their own career needs (36%) or both their own and the Air Force's needs (57%). Volunteers were more likely to cite bad working conditions, a boring job, or family concerns as reasons for leaving the earlier specialty.

Job Attitudes and Work Assignments

One approach used to evaluate whether job transfers under voluntary or involuntary circumstances had differential impacts was to compare volunteers' and selectees' self-reports of their job attitudes and work assignments. Perceptions of job satisfaction, perceived use of talents and training, and opportunity for assignment to responsible positions were assessed. On these items retrainees used a 5-point rating scale with poor to excellent options to describe their experiences in both their prior and current specialties.

Item means and standard deviations of ratings by the two retrainee groups of both AFSs are shown in Table 2. Two sets of statistical

contrasts were conducted to evaluate retraining impact. The first set addressed the question of whether volunteers' and selectees' attitudes and work assignments differed in either the previous or current AFS. The second evaluated whether there was any change in the experiences reported before and after the specialty reassignment. T-tests were conducted to evaluate the former (independent groups) and latter (correlated samples) research concerns. The resultant t-ratios and statistical significance decisions are shown in Table 3.

All contrasts of each of the three measures were statistically significant, and the direction of differences was the same. The representative data trend has been graphically depicted in Figure 1 to facilitate discussion of major findings. First, in the previous AFS selectees assigned a mean rating of "Average" to "Good" to their job satisfaction, use of talent and training, and responsibility level of work assignments. Volunteers' perceptions were significantly and appreciably (greater than one-half scale point) lower. Inspection of ratings in the current specialty revealed a reversal in the standing of the two groups. Volunteers' reports of their job attitudes and work assignments were substantially better than selectees'. The last major finding, as illustrated by the crossing lines, was that both the volunteers and selectees reported significantly different experiences in the two AFSs. Volunteers assigned more favorable ratings in the current AFS. However, selectees' descriptions of their job attitudes and work assignments were less positive in the current than in the former, before-retraining specialty.

Job Skill, Ability and Performance Assessment

Ability and performance on the job in the new specialty were examined using supervisors' ratings of retrained and non-retrained airmen on three appraisal factors. These were job skills and knowledge, supervisory skills, and quality of work. Item means and standard deviations were computed for the two retrained groups and for the non-retrained group (see Table 4). Other analyses were independent groups t-tests to determine if mean ratings assigned by the supervisors differed for selectees versus volunteers, selectees versus non-retrainees, and volunteers versus non-retrainees. T-ratios for the three comparisons are shown in Table 5.

Supervisors' appraisals of selectees, volunteers, and non-retrainees on the ability and performance items clustered, on the average, near a rating of "Good" (scale point 4) on the 5-point, poor-to-excellent rating scale. All statistical contrasts were significant, with the exception of selectees versus non-retrainees on the supervisory skills item. However, none of the differences in mean ratings was judged to be of practical significance. All were less than one-third scale point (and of a standard deviation unit, typically). Large sample sizes contributed to the extreme sensitivity of the statistical tests to small differences in mean ratings.

Several trends in the data were noteworthy. Selectees consistently received the lowest overall ratings on the skills and performance measures. The magnitudes of the rating differences between the selectees and the other two groups were typically greater than between the volunteers and non-retrainees. On the average, supervisors rated the voluntary retrainees and non-retrainees more similarly.

Adjustment to the Retraining Specialty

Supervisors' appraisals of attitudes toward work, motivation to do a good job, and interpersonal relations with co-workers were used as indicators of how well retrainees had adapted to the retraining experience and new occupation. Analysis procedures for the adjustment measures paralleled those for the performance criteria. Results are shown in the lower part of Tables 4 and 5.

The selectee, volunteer, and non-retrainee groups each received ratings which fell, on the average, near a rating of "Good" on the 5-point supervisory appraisal scale. Selectees received significantly lower ratings than did either the volunteers or the non-retrainees on the three adjustment measures. However, the magnitudes of the differences between the groups were small (usually less than .2 scale point) and were not considered to be appreciable. Statistical contrasts between volunteers and non-retrainees were not significant.

Selective Retraining Policy Issues

The views of both retrainees and supervisors on policy issues related to selective retraining concerns were solicited. Questions were designed to elicit information which retraining managers could use to restructure policies to provide a more favorable environment for selective retraining.

Retrainees were asked to judge what the overall impact of selective retraining would be on their productivity, motivation, and morale in the new job, and on their desire to remain in military service under three policy-related conditions. The conditions were selective retraining (1) without a choice of retraining AFS; (2) with several retraining AFSs from which to select; and (3) with choice of base of assignment in conjunction with retraining. Retrainees reported their opinions using a 5-point rating scale with end point descriptors which read "large negative" and "large positive" impact. Item means, standard deviations, and direction of results were similar for the four measures. Representative findings for the productivity item are shown in Figure 2. Selectee and volunteer groups were highly consistent in their judgments of the policy alternatives. Average ratings of impact of selective retraining on productivity were clearly negative without a choice of AFS. With a choice of several AFSs or base of assignment, the retrainees judged that their productivity would not be affected appreciably by selective retraining.

The supervisors' viewpoints were solicited on whether or not there should be a cut-off time for involuntarily retraining an enlistee out of his/her AFS and, if so, at what point in the military service career. The majority of supervisors (75%) favored a cut-off time not later than

the 15th year of service. The minority opinion (21%) was that there should be no time restriction on involuntary retraining. These supervisors' views were in concert with operational practice at the time of the survey.

Discussion and Conclusions

Collectively, the findings suggest that airmen whose retraining is compulsory or involuntary may be expected to experience more difficulty transitioning to their new specialties than volunteers. Selectees are apparently more resistant to the job change, as evidenced by their reports of poorer job attitudes and work experiences in the current specialty than in their former, before-retraining AFS. They were also less favorably disposed toward the current specialty than were their voluntarily retrained cohorts. Supervisors' assessments complemented the selectees' reports. Supervisors consistently gave selectees slightly lower marks on performance factors than volunteers. The supervisors seemed to be of the opinion though that both groups had acquired the requisite job skills and knowledge and were performing at satisfactory levels. Appraisals of work attitudes, motivation, and interpersonal relations on the job suggested that the selectees had not adjusted to the new occupational environment with the same ease as had volunteers. The volunteers, moreso than the selectees, seemed to be viewed by supervisors as performing and interacting on the job at a level comparable to that of non-retrained airmen.

Management factors were identified in the study to attenuate some of the observed negative impacts of selective retraining. Based on supervisors' reports, retraining managers have recently implemented a cut-off time at the 13-year point for retraining airmen involuntarily. The introduction of new opportunities for choices into the retraining decision system would also be expected to have a mitigating influence on the consequences of involuntary job transfers. The options evaluated in the study are not interpreted to be the best or only ones to make available to prospective selectees, but the retrainees' ratings do demonstrate the importance retrainees attach to having alternatives for consideration. As a whole, the current findings suggest that job changes made on a selective basis should be done with caution and used only to the extent needed to fulfill essential manpower requirements.

References

Air Force Regulation 39-4. Airman retraining program. Washington, D. C.: Department of the Air Force, 28 November 1979.

Miller, R. G., Jr. Simultaneous statistical inference. New York: McGraw Hill Book Company, 1966, 67-70.

Skinner, M. J. An evaluation of the Air Force airman retraining program. Proceedings of the 23rd Annual Conference of the Military Testing Association, Arlington, VA., October 1981.

Table 1. Percentage of Agreement/Disagreement Between Airmen Reports and Personnel Records on Volunteer/Selective Status

Personnel Record	Airman Self Report: Volunteer	Airman Self Report: Selectee
Volunteer	77%	18%
Selectee	2%	3%

Table 2. Voluntary and Selective Retrainees' Appraisals of Job Attitudes and Work Assignments in Previous and Current AFS

	Volunteer				Selectee			
	Previous AFS		Current AFS		Previous AFS		Current AFS	
Item	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Job Satisfaction	2.73	1.39	3.74	1.15	3.66	1.31	2.97	1.32
Use of Talents/Training	3.03	1.33	3.75	1.10	3.71	1.26	3.12	1.26
Opportunity for Responsible Work	3.01	1.37	3.82	1.20	3.63	1.33	3.33	1.35

Table 3. T-Ratios for Comparisons of Job Attitude and Work Assignment Ratings

	Volunteer vs Selectee		Previous vs Current AFS	
Item	Previous AFS	Current AFS	Volunteer	Selectee
Job Satisfaction	-31.12*	29.89*	-54.30*	17.80*
Use of Talents/Training	-23.55*	25.49*	-41.56*	16.27*
Opportunity for Responsible Work	-21.04*	18.06*	-44.33*	7.97*

* $p < .01$ (Bonferroni $\alpha = .01$).

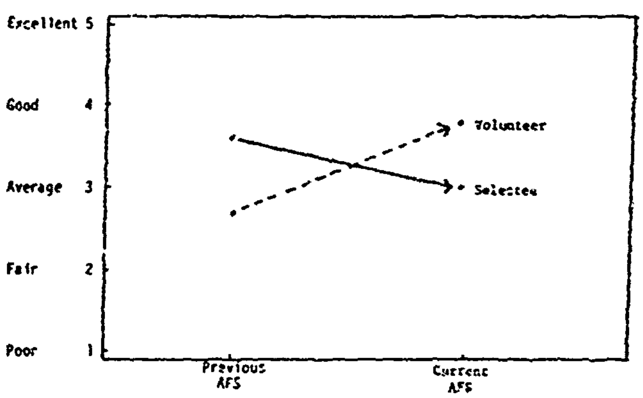


Figure 1. Representative Trend in Job Attitude and Work Assignment Ratings.

Table 4. Supervisors' Appraisals of Performance and Adjustment by Volunteers, Selectees, and Non-Retrainees

	Volunteer		Selectee		Non-Retrainee	
Item	Mean	S.D.	Mean	S.D.	Mean	S.D.
Performance						
Job Skills/Knowledge	3.91	1.07	3.75	1.07	4.12	.90
Supervisory Skills	3.66	1.12	3.50	1.18	3.55	1.10
Quality of Work	4.04	1.02	3.86	1.08	4.14	.93
Adjustment						
Attitude Toward Work	4.07	1.01	3.86	1.08	4.05	.95
Motivation/Good Job	4.13	1.01	3.96	1.09	4.15	.96
Co-Worker Relationships	4.06	.99	3.93	1.02	4.10	.91

* $p < .01$ (Bonferroni $\alpha = .01$).

Table 5. T-Ratios for Comparisons of Performance and Adjustment Ratings

Item	Selectee vs Volunteer	Selectee vs Non-Retrainee	Volunteer vs Non-Retrainee
Performance			
Job Skill/Knowledge	-6.91*	-15.76*	-12.28*
Supervisory Skills	-6.24*	-2.05	5.41*
Quality of Work	-7.61*	-11.67*	5.90*
Adjustment			
Attitude Toward Work	-8.82*	-7.78*	.37
Motivation/Good Job	-7.27*	-7.65*	-1.04
Co-Worker Relationships	-5.42*	-7.15*	-2.59

* $p < .01$ (Bonferroni $\alpha = .01$).

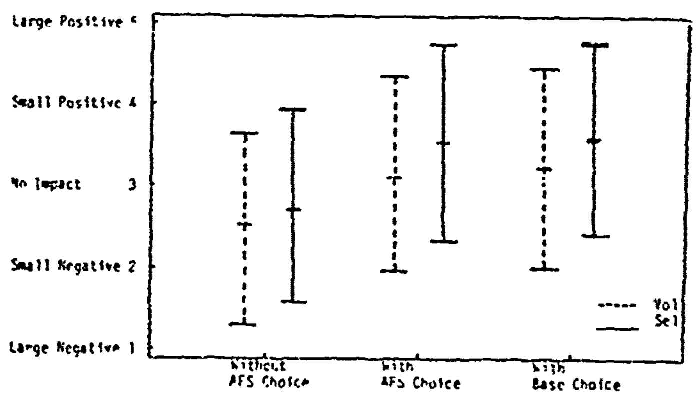


Figure 2. Means and Standard Deviations of Involuntary Retraining Impact Ratings on Productivity Under Three Policy Conditions.

Evaluating Individualized and Group Instruction Programs for Technical Associative Structure

BRANDON B. SMITH
Minnesota Research and Development Center
For Vocational Education
University of Minnesota
Minneapolis, Minnesota

In recent years, considerable amount of time and dollars have been expended to develop and implement individualized/computer assisted/managed technical programs in vocational education (Oen, 1973) and in the military (Brown and Rubininstein, 1976) (Brown, DeKleer, and Bernhain, 1976) (Tennyson, 1981) (King, 1975) (Kingsley and Slelzer, 1974) (Judd, O'Neal, and Spelt, 1974) (Dallman, and DeLeo, 1977). Textbooks also have been written on the topic of Individualized Instruction (Pucel and Knaak 1975), Howe, 1971). At least two studies have been conducted to evaluate individualized/computer assisted instruction (Orlansky and Streng, 1979, Dallman and DeLeo, 1977).

While it seems logical that as computer technology advances, it should be made use of for instructional purposes, the fact remains (Orlansky, and Streng, 1979) individualized instruction/computer assisted instruction (a) is effective for those who are able to complete the technical program, but it tends to produce more program attrition than traditional group instruction programs. Criteria used are (a) achievement measures, (b) task, performance and (c) field work competence.

Rationale/Purpose

The purpose of this presentation is to evaluate two individualized and group instruction tool and die programs in terms of the technical structure of associative knowledge and performance of experienced tool and die workers as compared with technical student of knowledge high and low ability students enrolled in two Minnesota Post-Secondary Vocational Tool and Die programs. It was reasoned that because individualized and/or computer assisted instruction programs are for the most part predicated on behaviorist principles (e.g., one single instructional frame presentation (a) reinforcement (b) feedback of results and (c) sequential accumulation of knowledge). from simple to complex. It is hypothesized that the learner may have difficulty integrating or assimilating this knowledge into a more global structure of technical knowledge. On the other hand, group instruction provides learners with the opportunity to discuss knowledges and skills with the instructor and peers and thus may have evolved a more integrated conceptual structure of knowledge leading to improved task performance.

Population/Sample

Five different tool and die workers in the Minneapolis area were contacted and empoyers asked to independently rate all their tool and die workers on several criteria (1) work variety, (2) versatility/adaptability,

(3) creativity problem solver, (4) most accurate, and (5) efficiency, (6) quality. From these ratings, one worker from each of the five firms was selected to participate in the study.

Similarly, post secondary vocational instructors rated each student in their respective individualized and group instruction program as to the students ability to learn the content and successfully perform the tasks in a 2 year tool and die program. Based on teacher ratings of all students, two groups of five high and low performing students in the individualized and group instruction program respectively were identified to participate in the study.

Methodology/Administration

The free association methodology was used to identify and compare the associative conceptual structure of knowledge for a purposive simple of five groups of individuals in this study. The rationale for the methodology is based on the previous work of various verbal learning theorists (Deese, 1962), Garskoff and Houston, 1963),Johnson (1964), 1967) and the previous work of (Smith, 1968), Pratzner (1970), (Liu, 1972), (Nee, 1978) (Ammerman, (1970) for application in vocational education and the military fields.

The rationale for the free association methodology suggests that technical workers possess verbal labels for the technical concepts in their field. By obtaining free association responses from them for a population of technical stimulus words in their technical field it is possible to determine the meaning of these words and then generate the asssociative structure of knowledge of the words. The relationship among the associative menaning of these technical words, will form a hierarchical association structure which can be used to evaluate group and individualized technical instruction programs.

The free associative methodology as is based on several principles or assumptions about the verbal behavior of individuals in various technical fields.

1. All technical fields used technical words to communicate and teach technical concepts.
2. Workers and students use these technical terms to communicate and understand these concepts.
3. Workers and students are capable of responding to a stimulus word with relevant technical responses.
4. Workers and students organize their technical concepts into an integrated structure dependent upon their functional relationship to their work role/technical learning environment.
5. Relatedness coefficients can be computed for all possible combinations of stimulus word response and subjected to higher order factor analysis to generate a hierarchical associative technical structure of knowledge for a group of individuals who are known to possess and be performing at qualitatively different levels.

A sample of 85 technical stimulus words were selected and administered to the five different groups in two test sessions where each group was asked to respond to each stimulus word with as many relevant technical words they could think of in a one minute time period. The associative technical meaning were obtained for each of the 85 technical terms and for each of the respective five groups of individuals by creating a rank ordered distribution of all the responses given at least two times by the respective groups (pooled associative meaning). A relatedness coefficient (RC) matrix was generated which computed the amount of relationship among all possible combinations of stimulus word response distributions and then subjected to a hierarchical factor analysis procedure to produce a hierarchical conceptual associative structure of knowledge for each of the five groups. This provided the opportunity to compare both the verbal behavior of the five groups as well as the graphic structure of how these concepts were differentially integrated by the five groups.

Objectives

The purpose of this study was to evaluate and compare the five different groups in terms of (1) test-retest reliability or their free response (2) the size of the technical vocabulary (3) the number of factors in the heirarchical associative technical structure and (4) the relationshp of the associative structure to a performance task.

Test-Retest Reliability of Responses

Table 1 shows the test-retest reliability coefficents for each of the five different groups for a random sample of fiteen different stimulus wards. The coefficient of stability range from a low of .38 to a high of 1.00 for each of the fifteen words with an average stability coefficient of about .80. This tends to indicate that the verbal responses of the workers and high and low ability students are quite reliable and thus are capable of producing a stable associate technical structure of knowledge.

The associative technical structure of knowledge is a function of the size of the technical vocabulary of a individual or group of individuals. It may be hypothesized that workers would have the largest technical vocabulary and the most integrated structure of technical concepts followed by high ability students in group instruction and high ability students in individualized instruction programs. Low ability students would have the smallest technical vocabulary and the least integrated structure of knowledge.

As can be seen in table 2, workers have the largest technical vocabulary, the largest pooled technical vocabulary and used the fewest different words. On the other hand, low ability students in the individualized instruction program had the smallest pooled technical vocabulary followed by the low ability students in the group instruction programs. While the differences were not great, students enrolled in group instruction program (both high and low ability) seem to have a larger, more agreed upon technical vocabulary then either the high or low ability students in the individualized instruction program.

Performance

Table 3 shows the correlation of the rankings of the five groups of individuals in a cognitive performance task believed to be related to their total understanding of the tool and die field. Each group was given a tool and die part as a sample and a drawing of the part and were given a list of eighty randomly ordered sstatement necessary to design, plan and make the part. The correlations among this ranking indicate relatively low correlations of the worker group with any of the other student groups. The highest correlation was between (1) high and low ability students enrolled in group instruction (.595) and (2) between high group instruction students and high individualized instruction students (r=.515) the lowest correlations were between low ability individualized instruction students and high ability individualized instruction students (-r=.011).

The general conclusion which seems most is plausable is that in terms of the correlations among ranking of tasks. (1) high ability students are more in agreement on tasks with worker rating then low ability students, (2) high abiilty students in both instructional programs tend to relate tasks similarly and (3) ther is a negative and low agreement among high and low ability students who are envolved in individualized instruction then for high and low ability students envolved in group instruction group.

Structure of Knowledge

The hierarchical function analyses of the associative structure of knowledge has at this time not been completed, but will be completed by the time the report will be made. First, order factor analysis has been completed and the results are as follows:

Workers	32 Factors	Group Inst. High	34 Factors
Group Inst. Low	29 Factors	Ind. Inst. High	31 Factors
Ind. Inst. Low	33 Factors		

Conclusion

Preliminary conclusion suggest the following:

1. Free association responses were quite reliable for all groups.
2. Workers tend to have a larger more consistent technical vocabulary then either high or low ability students in either group or individualized technical programs.
3. High ability students regardless of the mode of instruction tend to perform better than low ability students, but low ability group instruction students seem to perform as well as high ability students in either group or individualized instruction. Low ability in individualized instruction programs do least well in performance.
4. It is anticipated the hierarchical factor analysis of the five different groups will also demonstrate quantative and qualitative differences among the groups in terms of (a) the number of factors, (b) the number of levels in the hierarchical associative structure, (c) the integration of the structure and (d) the tables for the factors.

References

Ammerman, H. L., "Systematic Approaches for Identifying and Organizing Content for Training Programs," Professional Paper, Human Resources Research Organization, Alexandria, Virginia, 1970.

Brown, J. S. et. al. Intelligent Computer Assisted Instruction (CIA) Applications. DHEL-TR-76-67, Technical Training Division. Lowry Air Force Base, Colorago 1976.

Brown, J. S. et. al. Reactive Learning Environment for Computer Assisted Electronics instruction, AFHRL-TR-76-88 Technical Training Division Lowry Air Force Base, Colorado, 1976.

Dallman, B. E. and Delea, P. J. Evaluations of Plato IV in Vehicle Maintenance Training, AFHRL-TR-59. Technical Training Division, Lowry Air Force Base, Colorado, 1976.

Deese, J. "On the Structure of Associative Meaning", Psychological Review 69: 161-75, 1962.

Deese, J. The Structure of Assciations in Language and Thought, Baltimore: The Johns Hopkins Press, 1965.

Garshof, B. D. and Houston, J. P. "Measurement of Verbal Relatedness: An Ideographic Approach," Psychological Review 70:377-88, 1963.

Howes, Virgil M. Individualization of Instruction: A Teaching Strategy. The MacMillin Company, New York, 1971.

Johnson, P. E. "Associative Meaning of Concepts in Physics," Journal of Educational Psychology 55:84-88, 1964

Johnson, P. E. "Some Psychological Aspects of Subject Matter Structures," Journal of Educational Psychology, 56: 217-224, 1965.

Judd, W. A. et. al. Individual Differences and Learner Control: Investigation of Control Over Pictorial Mediators in Computer Assisted Instruction, AFHRL-TR-73-30, Technical Training Division, Lowry Air Force Base, Colorago, 1974.

King, A. T. Impact of Computer Based Instruction on Attitudes of Students and Instructors: A Review, AFHRL-TR-75-4. Technical Training Division, Lowry Air Force Base, Colorado, 1975.

Kingsley, E. J. and Stelzer, J. The Theoretical Bases for Individualized instruction, AFHRL-TR-74-10. Human Resources Research Office, Alexandria Virginia, [illegible]73.

Lui, C. C. "Testing an Empirical Procedure for Indentifying Technical Associative Structures, Discrimiiniating Between Students Grouped by Performance and by Creative Thinking in a Mobile Communication Technology Program," Unpublished Doctoral Thesis, Minneapolis, Minnesota, University of Minnesota, 1972.

Nee, G. "The Development of a Technical Conceptual Structure for the Concepts Possessed by Selected Quality Control Specialist," Dept. of Industrial Education Technology, Central Michigan University, Mount Pleasant, Michigan, 1977.

Oen, U. T. Integrating the Interaction of Learning Styles and Type of Learning Experiences Vocational Technical Education, Office of Education, Washington, D.C., 1973.

Orlanski, J. and String, J. Cost-Effectiveness of Computer-Based Instruction in Military Training, IDA Paper P-1375, Institute for Defense Analysis, Science and Technology Division, Arlington Virginia, 1979.

Pratzner, F. C. "Testing an Empirical Procedure for Identifying Technical Associative Conceptual Structure: Discriminating Between Workers Within and Between Two Occupations," Unpublished Doctoral Thesis, Minneapolis, Minnesota, University of Minnesota, 1969.

Pucel, D. J. and Knaak, W. Individualizing Vocational and Technical Instruction. Merrill, Columbus, OH 1975.

Smith, B. S. and Moss, J. (Ed), "Report of a Seminar: Process and Techniques of Vocational Curriculum Development," Minneapolis, Minnesota, Minnesota Research Coordinating Unit for Vocational Education, 1970.

Smith, B. S., "Testing an Empirical Procedure for Identifying Technical Associative Conceptual Structure: Discriminating Between Flexible and Inflexible Radio and Television Repairmen," Unpublished Doctoral Thesis: Minneapolis, Minnesota, University of Minnesota, 1968.

Tennyson R. D. "Use of Adaptive Information for Advisement in Learning Concepts and Rules Using Computer Assessted Instruction", American Educational Research Journal, Vol. 18, No. 4, 1981.

Management Information System for Processing USCG Class "A" School Requests

by James R. Stokes
Psychological Research Branch, Office of Personnel, U.S. Coast Guard

Many managers today require immediate access to accurate data for generatation of timely output. Such was the case of the Training and Education division of the Office of Personnel, U.S. Coast Guard. My branch, Psychological Research, used a powerful, interactive language called APL that enabled us to design and implement a management information system for procssing school request applications. Although this system was specifically for training involving schools, it could just as well been used for any other type of desired information.

For years the Training and Education Division of the Coast Guard had a problem. They lacked timely access to information on non-rated enlisted personnel applying for a speciality school (i.e. Boatswain's Mate School, Aviator Technician -- the Coast Guard refers to these schools as "A" schools). A computer list from the Department of Transportation's computer center was only provided every six weeks and contained many mistakes. This outdated and inaccurate data caused many headaches for the Training and Education Division. It was not unusual for twenty-five messages and forty phone calls to be received in one day from frustated applicants who really did not know where they stood on a particular school list. In one instance, a thoroughly disgusted chief sent fifty copies of his "dream sheet" application to the division by registered mail. The lack of up-to-date lists were the cause of many congressional inquiries.

The Psychological Research Branch, Office of Personnel was approached by the Training and Education Division for help. After a series of meetings discussing the "A" school list problem, it was decided that a management information system was the solution. Ideally, the Training and Education Division wanted an internally controlled system that would allow for input, modification, removal and listing of applicant's requests. Later, if feasible, the information system could provide direct printout of orders or assist somehow in the process of creating orders.

The Training and Education Division conferred with various local contractors to see how much the cost of such a system would be. When they were told that the price would be in the neighborhood of $130,000, the division's representatives turned to the Psychological Research Branch to see if the task could be done in-house. After a series of meetings to determine just exactly what was needed, Psychological Research agreed to take on the project. The project was named TMIS (Training Management Information System).

This project was a tremendous success. The Training and Education Division now has real-time access to files for input, correction, and removal of an applicant's school request. Immediate output allows the division to check for any input mistakes and make corrections on the spot. By mailing copies of the up-to-date lists to different districts, the Training and Education Division no longer spends a majority of the day answering questions and handling complaints. Writing orders to a particular school is both simpler and quicker. Most important, TMIS provided the Training and Education division 98% of what they had originally wanted from a contractor, and saved the Coast Guard over $100,000.

The first stage of TMIS is the INPUT PHASE. To begin entering applicants' information into the data base, the user types in "SCHOOLINP" and the computer responds, asking for information. Figure 1 shows the terminal input by the user and correspoding computer responses.

Figure 1 INPUT PHASE

```
User:       SCHOOLINP

Computer:   NAME
User:       SMITH ZZ

Computer:   SSN
User:       123456789

Computer:   EOS
User:       841212

Computer:   CO WAIVER  (YES OR NO)
User:       YES

Computer:   GCT
User:       67

Computer:   ARI
User:       65

Computer:   MECH
User:       68

Computer:   ETST
User:       65

Computer:   CLER
User:       62

Computer:   1ST PREF 'A' SCHOOL
            SCHOOL1
User:       AD

Computer:   PHYS REC'D  (YES OR NO)
User:       YES

Computer:   PASSED OR FAILED
            (ENTER "P" OR "F")
User:       F

Computer:   2ND PREF 'A' SCHOOL
            SCHOOL 2
User:       YN

Computer:   3RD PREF 'A' SCHOOL
            SCHOOL3
User:       XX

Computer:   PAYGRADE
User:       3

Computer:   MINORITY
User:       5

Computer:   SEX
User:       1

Computer:   SIGDATE
User:       800820

Computer:   OPFAC7
User:       1730520

Computer:   UNITREPDATE
User:       801009

Computer:   ANOTHER ENTRY?
User:       NO
```

All input is checked for at least some degree of validity. For example, if the user enters a letter when the computer had asked for social security number, the user would get an error message. When an aviation school is entered by the user, the system executes a subprogram to request information on the status of the applicant's physical. When a unit identification number (OPFAC) is entered, the system goes to a separate data base to extract the correct unit mailing address. When input is completed, all information is stored in a data base.

The second stage of TMIS is the MODIFICATION PHASE. This phase allows the user to correct information on an applicant's request, and is shown in Figure 2.

Figure 2 MODIFICATION PHASE

User: MOD

Computer: ENTER NAME OF APPLICANT

User: POWERS TE

Computer: HERE ARE THE FIELDS AND CURRENT VALUES FOR: POWERS TE:

FIELD	CURRENT VALUE
NAME:	POWERS TE
SSN:	123123123
EOS:	841212
GCT:	67
ARI:	63
MECH:	67
ETST:	70
CLER:	58
SCHOOL1:	ET
SCHOOL2:	SK
SCHOOL3:	XX
PAYGRADE:	3
MINORITY:	5
SEX:	1
SIGDATE:	800202
OPFAC7:	1730520
UNITREPDATE:	801009
PHYS:	
CO:	5

Computer: ENTER THE FIELD YOU WISH TO CHANGE

User: PAYGRADE

Computer: ENTER NEW VALUE FOR PAYGRADE

User: 4

Computer: ANOTHER CHANGE?

User: NO

After modification is complete, the new updated information is written out to the data base.

The third stage of TMIS is the REMOVE PHASE. It is executed when the user wishes to remove an applicant request from the data base. The appropriate commands and corresponding responses are shown in Figure 3.

Figure 3 REMOVE PHASE

User: REMNAME

Computer: ENTER NAME OF APPLICANT
User: POWERS TE

Computer: DID THIS PERSON GO TO 'A' SCHOOL?
User: NO

Computer: ANOTHER ENTRY?
User: NO

At this point, T. E. Powers' application for 'A' school is removed from the data base.

The fourth stage of TMIS is the LIST PHASE. When the user wishes to see a list of applicants' by a particular school, he enters the command "LIST" on the computer terminal, as shown in Figure 4. The list is prioritized by signature date of the application, and paygrade.

Figure 4 LIST PHASE

User: LIST

Computer: SCHOOL
User: AD

Computer: BY A PARTICULAR DISTRICT?
User: YES

Computer: DISTRICT
User: 01

Figure 5 illustrates the resulting ouput from the LIST phase.

Figure 5 Output from LIST PHASE

3:45 PM EDT 5/19/81

AD LIST

DISTRICT 01

NUM	TEST	PAYGR	PHYS	NAME	SCHOOLS	UNITREF DATE	CONUS	EXP DATE	TIS	SIG DATE	UNIT
15		3		DELAVRO MJ	AD XX XX	800624		840420		801213	BASE SOUTH PORTLAND OR
17		3		JEZIERSKI	AD XX XX	800311		841212		801215	COGARD TRACEN CAPE MAY NJ
20		3		MARZULLA MC	AD XX XX	800627		841212		801217	USCGC ACTIVE
25		3		PORAZZO PJ	AD XX XX	800305		841212		810115	USCGC UNIMAK
30		3		DAVIS RK	AD XX XX	800909		840701		810215	COGARD BASE HONOLULU HI
37		3		MOORE RL	AD XX XX	800910		841212		810316	USCGC BUTTONWOOD
43		2		ALBEE FB	AD XX XX	800717		840309		800130	USCGC BIBB
44		2		INGHRAM DR	AD XX XX	800202		841212		800201	USCGC UNIMAK
58		2		BOYNTON GH	AD XX XX	800310		831025	N	800811	USCGC CHASE
75		2	F	DELEO CM	AD XX XX	801028		840824		801123	USCGC POLAR SEA
85		2		NOLDER D	AD XX XX	801208	I	8[illegible]1005		801217	USCGC CAPE HORN

Internally, TMIS selects from the data base only those applicants who have applied for that school. It then sorts them by paygrade and signature date, and assigns them a priority number. At this point, if the user has requested only a particular district, the computer selects only that information. It then checks time in service, test scores for that school, time served inside the continental United States (INCONUS) or outside (OUTCONUS), and for physical status (aviation schools only). Applicants must meet these four qualifying factors in order to become eligible for school. The system then generates internal "flags" for non-qualified candidates. It then formats the data and lists the information at the user's terminal. TMIS also contains a method that allows the user to ouput the list to a laser printer.

Finally, there is the ORDERS PHASE. This feature allows the user to create order variables on selected applicants. The information is then transferred to a word processor where orders can be written. These particular applicants are then removed from the main data base and placed in another file which contains information on applicants sent to school.

The TMIS system is now in place and fulfilled over 95% of a contractor's proposal at a fraction of the cost. The flood of telephone calls and other problems were reduced considerably because of the system. TMIS lets managers quickly process applications. At present, the system is undergoing expansion.

If other managers have similar problems regarding training -- or, for that matter, any other type of information storage and retrieval problem --, I would recommend considering a system similar to the one implemented by our branch in the computer language APL.

APL LANGUAGE

The Psychological Research Branch uses the computer language APL for a variety of purposes. In addition to this type of management information system, APL is useful and efficient in areas such as simulation modelling and budget reports. One reason our branch agreed to tackle TMIS was that APL is a language highly suited for information projects of this type. Unlike most other languages, APL is interactive to begin with, and thus fits in smoothly with the interactive requirements requested by the Training and Education Division. Coding and executing APL subprograms directly at the terminal cuts down on programming and debugging time. Inside the language, Boolean operators efficiently allow for capturing and removing data without looping. Also, the language uses symbols instead of words to execute commands. This reduces the number of lines of code and, ultimately, programming time.

Here are some the symbols used in APL:

/ \ ρ ~ ↑ ↓ ○ ⋆ [⌈ ⌊ _ ∇ ∆ ∘ ' ⎕ ← → ∩] ⊥ ⊤ | ( ) - + × ÷

Finally, in APL there is no compiling and loading, just executing. It is almost like working in the load module itself. By using APL in place of other languages for applicable projects, an analyst may well find that system design is easier, less time is spent programming, debugging is faster, and additions and modifications become much simpler.

AD P000912

Job Difficulty Data as an Indicator of Job Complexity

Thomas E. Ulrich and Charles D. Gorman
United States Air Force Academy

It is widely acknowledged that the technological advances of the present are going to have a major impact on the design and operation of future weapons systems. Future systems will continue to depend on the performance of the operator or the maintenance person assigned responsibility for the system.

Addressing this multifaceted problem calls for a focus on the attitudes of the people involved with respect to increasing complexity. It is clear that these attitudes can impact many elements of the system, from personnel manning to performance levels. The investigation of this area prompted the application of the Job Difficulty data obtained by the U.S.A.F. Occupational Measurement Center (USAFOMC) in a way which had not been used before. In addition to providing data applicable to the Department of Defense (DoD) question of complexity, additional light was shed on the construct of job complexity as measured by the concept of job difficulty.

The Construct of Complexity

Complexity doesn't have a generally accepted operational definition. It has been measured as workload (Wierwille, 1979); ambiguity (Abdel-Halim, 1981); and stress (Chiles & Alluisi, 1979), among other labels. It is difficult to define the continuum of complexity, or pinpoint what makes one system more complex than another.

Hackman and Oldham's (1980) instrument called the Job Diagnostic Survey (the JDS) measures among other things five core job characteristics: skill variety, task identity, task significance, autonomy, and feedback from the job itself. These characteristics are measured via incumbents' responses to questions concerning the degree to which certain factors are present in their jobs. A measure of "job enrichment" is usually calculated by combining the five key job characteristics into a Motivating Potential Score (MPS) for the job in question.

The five core job dimensions have been used as a measure of job complexity (Abdel-Halim, 1978; Katerburg, Hom, &

Hulin, 1978). Low scores on the job complexity measure are, according to Hackman and Oldham (1980), descriptive of simple, structured jobs while high scores are descriptive of relatively complex, unstructured jobs. Dunham (1977) has demonstrated that jobs high in complexity have higher job ability requirements, and that job complexity is positively related to job ability requirements.

Subjectivity in Complexity Measurement

There has been widespread use of job complexity scores based on the Hackman and Oldham instrument as indicators of task content (Roberts & Glick, 1981), yet there has been relatively little work outside the human factors literature which has looked at the complexity of jobs from the task structure viewpoint. The measurement via the JDS of work performed by incumbents is actually a measure of perceived task and job complexity. O'Reilly (1977) has criticized the subjective perceptual nature of the measurement of job characteristics. Perceptual measures of task design confound individual differences in perception with the objective task characteristics. In response to this, investigators have called for more objective measurement of job characteristics (Pritchard & Peters, 1974; Roberts & Glick, 1981)

The Use of Job Difficulty Data

A more objective measure of job complexity can be found in the USAFOMC measure of job difficulty. For each enlisted specialty surveyed, ratings are obtained from experienced specialists in the career field as to the difficulty level of each task which appears in the job task inventory. Difficulty is defined as the amount of time needed to learn to do the task. Selection of the USAFOMC measure for this purpose represents a new application of the job difficulty data.

Method

Sample Selection

The first stage in the investigation of the relationship between complexity (as measured by job difficulty) and job attitudes was a review of current occupational survey reports (OSRs) with personnel from USAFOMC. It was important to find a career field which had been recently surveyed, as well as one with job groups ranging from very difficult to very simple jobs. The specialty selected was the

325X0 career field, Automatic Flight Control Systems (AFCS). Within that specialty, four groups were selected; the groups are presented in Table 1.

Table 1

Job Groups Selected

	Group 297	Group 274	Group 396	Group 581
Group Title:	TAC AFCS Flightline Personnel	TAC AFCS Shop Personnel	MAC C-141 Flightline & Shop	MAC C-5 / C-141 Flightline & Shop
Job Group Size: (OSR)	n = 45	n = 39	n = 35	n = 38
Job Difficulty Index:	7.00	9.44	11.53	16.55
* Number of incumbents surveyed:	35	45	36	74

* The size of the sample surveyed exceeded the number of original job group incumbents at bases where all AFCS shift personnel performing the same job were contacted.

Data Collection

Once the groups were chosen, the PRTVAR listing of respondents was obtained for the individuals within each of the job groups. The next step involved contacting the organizations at each base from which incumbents in the selected job groups had come, and arranging in-person or by mail administration of the survey instrument. At each base, an attempt was made to insure that we were targeting the same job group which had originally made up the job group in the OSR. A check was also made with the NCOICs of the maintenance shops to verify that the tasks which were cited in the OSR as distinctive for the selected job groups were indeed characteristic of the job the current incumbents were performing. Survey data were collected from 190 325X0 personnel in this manner.

Characteristics of the Complexity Level Groups

The validity of selecting the job difficulty measure as an objective measure of complexity depends on the accuracy of the difficulty score obtained in the occupational survey as it applies to the incumbents sampled in the current study. When the surveys were received and the initial data was examined, it became clear that one of the job groups (Group 274) was appearing as performing more complex work than would have been expected. In fact, the complexity level (and other measures) for that group, which should have

been second to lowest, was higher than any of the other three job groups on the measures of interest. The findings for these measures are presented in Table 2. The initial interpretation was that the difficulty and complexity measures were thus tapping different constructs; however further examination revealed another possible explanation.

Table 2

Difficulty and Complexity Measures

	Group 297	Group 274	Group 396	Group 581
Job Difficulty Index:	7.00	9.44	11.53	16.55
Motivating Potential Score:	120	129	117	127
JDS Job Complexity:	80	87	83	85

Review of the background data (equipment worked on, systems maintained, etc.) revealed that the incumbents who had been selected as representative of Job Group 274 were reporting working on many more different systems than did the original sample from the job group. Bases surveyed by mail were contacted by phone and asked about the systems worked on. It was confirmed that the data we received were correct; the jobs had changed in the time since the job inventory had been administered, and additional aircraft were now part of the responsibility of incumbents at several of the bases.

Examination of the other job groups did not reveal any similar major discrepancy, so analysis continued with the three remaining groups (and data from Group 274 was withheld from further analysis). It would not have made sense to proceed with Group 274, since the job difficulty level of that sample was no longer known.

Objective vs. Subjective Measures of Complexity

The measure of job difficulty was compared with the Hackman and Oldham measure of job complexity For the three job groups used in the analysis, the correlation between

difficulty level and JDS complexity was only +.13, which was significant only at the $p < .10$ level. Thus the relationship between the objective measure and the JDS measure of complexity was not especially meaningful.

One key factor may be a clue as to why such a low correlation was observed. Based on the observation about job change within Group 274, it was clear that for that group at least the job performed at the time the job difficulty data had been collected had changed for many of the job group members. In the case of Group 274, many additional aircraft and thus additional autopilot systems had been added to the responsibilities of the incumbents in that job group, so the JDS measures were clearly describing different jobs than did the earlier job difficulty measure. We suspect that the time between difficulty and JDS measurement may have had a similar, though less obvious, effect on the other job groups which were included in the analysis.

Suggestions for Further Research

This problem of the changing of jobs over time could be eliminated if the job difficulty data and JDS data were to be collected at the same time. A future direction which will be pursued will be to have the two measures collected in a single administration of survey materials. The objectivity of the job difficulty measure will be retained because the actual computation of the difficulty level of the individual's job will be based on the independent evaluations of task difficulty. However, we will have the control of knowing that the incumbent's report of the task characteristics present in his or her job are measures of the same tasks and job on which we have a difficulty rating. The correlation between difficulty and complexity measures obtained in this way will be a much more reliable measurement of the relationship between the two variables.

Recent research on the aptitude requirements of Air Force jobs performed by personnel at the Air Force Human Resources Laboratory (Weeks, 1981) has included the development of "Benchmark Scales" for the measurement of job difficulty across specialties within an aptitude area. This methodology provides additional promise for the study of job complexity, by eliminating the limitations inherent in the within-specialty measurement of job difficulty used in the present study.

NOTE: This study was partially funded by Directorate of Programs, Analysis, and Evaluation (Tactical Air) under the office of the Secretary of Defense. Opinions expressed are those of the authors, and do not necessarily reflect DoD or USAF policies.

References

Abdel-Halim, A. A. Employee affective responses to organizational stress: Moderating effects of job characteristics. Personnel Psychology, 1978, 31, 561-579.

Abdel-Halim, A. A. Personality and task moderators of subordinate responses to perceived leader behavior. Human Relations, 1981, 34, 73-88.

Chiles, W. D., & Alluisi, E. A. On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 1979, 21, 515-528.

Dunham, R. B. Relationships of perceived job design characteristics to job ability requirements and job value. Journal of Applied Psychology, 1977, 62, 760-763.

Hackman, J. R., & Oldham, G. R. Work redesign. Reading, Mass: Addison-Wesley, 1980.

Katerburg, R., Hom, P. W., & Hulin, C. L. Relationships between job complexity and responses by organizational members: A test of moderating effects. (NR 170-802). University of Illinois Technical Report, 78-1, May 1978.

O'Reilly, C. A. Personality-job fit: implications for individual attitudes and performance. Organizational Behavior and Human Performance, 1977, 18, 36-46.

Pritchard, R. D., & Peters, L. H. Job duties and job interests as predictors of intrinsic and extrinsic satisfaction. Organizational Behavior and Human Performance, 1974, 12, 315-330.

Roberts, K. H., & Glick, W. The job characteristics approach to task design: A critical review. Journal of Applied Psychology, 1981, 66, 193-217.

Weeks, J. L. The development and application of measures of occupational learning difficulty Brooks Air Force Base, Tex.: Manpower and Personnel Division, Air Force Human Relations Laboratory (AFSC), 1981.

Wierwille, W. W. Physiological measures of aircrew workload. Human Factors, 1979, 21, 575-593.

Benchmarking Occupational Survey Task Factor Data

David S. Vaughan
University of Texas at Austin

One important type of data gathered by the occupational survey process is that on task factors. Task factors are task characteristics that are (at least relatively) independent of the particular jobs containing the tasks. Task factors routinely gathered by the U. S. A. F. Occupational Measurement Center include task learning difficulty and recommended training emphasis. Other task factors which have been gathered include consequences of inadequate performance and task delay tolerance.

Usually, such task factor data are gathered by having subject-matter experts--people who are familiar with most or all tasks in a particular Air Force Specialty (AFS) or occupation--rate tasks on the degree to which such tasks have the characteristic under consideration. In the Air Force, such subject matter experts are usually senior non-commissioned officers in the specialty being studied. Usually, task ratings are gathered on a relative scale, and the subject-matter experts are able to rate tasks in only one specialty.

Data gathered in this way allow tasks to be compared for the relevant characteristic (say, difficulty) <u>within</u> one specialty. However, it is not clear that ratings of tasks in one specialty can be compared with ratings of tasks in other specialties. For example, it is not clear that a task whose difficulty rating in one specialty was "five" has the same difficulty as a "five" task in a different specialty. This is the <u>benchmark</u> problem.

For certain practical applications, a solution to the benchmarking problem in task factor data is necessary. For example, we want to set aptitude requirements to enter various job specialties so that the specialties that are most difficult have the highest aptitude requirements. In order to use occupational survey data for the purpose of setting aptitude requirements, data on the task learning difficulty factor must be comparable across specialties.

In order to set aptitude requirements based on occupational survey data, the Air Force has a major research effort (Burtch, Lipscomb, & Wissman, 1982; Weeks, 1981). The approach used in this line of research involved finding and training subject-matter experts who could rate difficulties of tasks in many specialties. In addition, rating scales were constructed on which each point was "benchmarked" by several tasks which defined that particular point of the scale.

While this research effort has been successful in that task difficulty data have been gathered which are comparable across specialties, this

approach has proven to be very expensive and time-comsuming. The purpose of the present paper is to present a different approach for benchmarking task factor data. This approach is primarily statistical in nature. Task ratings on the factor to be benchmarked are gathered in the conventional manner; other data about the tasks are used to adjust the ratings onto a scale which is common to several specialties.

First, the statistical model of the proposed benchmarking will be presented. This will be followed by results of this method's application to some real task factor data.

Statistical Model

Consider task ratings of a factor in several different specialties. We will assume that a common "benchmarked" scale exists for the factor, although the observed ratings may not be on that common scale. Instead, we assume that the observed ratings in a specialty i are on an arbitrary linear transformation of the common scale, as

$$y(ij) = a(i)\, z(ij) + c(i), \qquad (1)$$

where
- y(ij) = observed rating on task j in specialty i,
- z(ij) = (unobserved) rating of task j in specialty i on the common "benchmarked" scale,
- and a(i), c(i) are constants for specialty i.

We further assume that ratings on the common scale, z(ij)'s, can be expressed as a linear combination of scores on some predictors, where this linear combination is the same for all specialties:

$$z(ij) = b'\, x(ij) \qquad (2)$$

where
- x(ij) is a vector of scores on predictor variables, for task j in specialty i,
- b is a vector of constant weights,
- and z(ij) is as defined as above.

We could use the a(i)'s and c(i)'s of equation 1 or the b-vector and predictors of equation 2 to obtain estimates of the task factor as measured on the common "benchmarked" scale. Since we do not know the z(ij)'s, we cannot solve either equation 1 or equation 2 for the desired values. But by substituting equation 1 into equation 2 we obtain an equation that can be used to estimate the needed values:

$$y(ij) = a(i)\,(b'\, x(ij)) + c(i). \qquad (3)$$

How can we go about estimating the parameters of equation 3 from real data? Probably the most straightforward approach is least-squares estimation--find values of the parameters so that the sum of squared devi-

ations between actual y-values and those predicted by equation 3 is as small as possible. Methods for least-squared estimation of linear equations are well-known. However, equation 3 is not linear in its parameters, since it involves products of the a(i)'s and the b's. Conventional least-squares methods cannot be used. However, general-purpose numerical optimization methods can be used for least-squares estimation of equations like equation 3, and that approach is used here.

Statistical inference procedures use with least-squares estimation of linear equations cannot, strictly speaking, be used for nonlinear equations such as equation 3. However, simulation studies (Duncan, 1978; Fox, Hinkley, and Larntz, 1980) have shown that, in practice, statistical inference procedures for least-squares estimation in linear models work reasonably well when applied to nonlinear (in parameters) models. Thus, that approach is used here.

An Example: Task Strength Ratings

In the previous section, a statistical model was presented for estimating benchmarked task factor ratings from relative ratings and other data. Here, an example will be given in which the model of equation 3 was fitted to real task factor data.

In this example, the task factor ratings to be benchmarked are ratings of the overall physical strength required to perform tasks. These strength ratings were gathered in the conventional manner. In addition, data were gathered on a number of variables which might predict overall strength. These data were gathered for approximately 25 tasks in each of eight job specialties (AFSs). The tasks and specialties were selected because they were thought to be likely to have significant physical strength requirements. The strength ratings were gathered from several subject-matter experts in each specialty; acceptable levels of interrater aggrement were obtained in each of the eight specialties. Details of the data-gathering and interrater agreement analyses are presented by Gott (Note 1).

The predictor variables (the x(ij)'s in equation 3) are summarized in Table 1.

Table 1

Predictor Variables

Type of work (lifting or lowering; 1 or 2 hands)
Amount of Repitition
Rate
Weight handled
Body posture (standing, sitting, crawling,
lying, kneeling, stooping,
bending at waist, swimming)
Position (distance above or below surface)
Altitude
Distance
Holding time
Time
Percent performing
Percent time spent
Environment (% indoors, outdoors, in flight)
Manpower required
Frequency (times per day, week, and month)

Many of the predictors listed in Table 1 are categorical variables. Such predictors were represented in the statistical model by sets on nonredundent variables indicating which categories were present for particular tasks. As a result, a total of 35 predictor variables were used in the analyses. Least-squares estimation procedures were used, as described above. All 35 predictors were used in all analyses. Up to 16 specialty parameters (2 parameters each for eight specialties) were estimated, based on 250 tasks. Data were available for 25 tasks each in six specialties and for 50 tasks in two specialties.

Three models were fit. One was the full model of equation 3. In addition, models were fit in which all the a(i)'s were constrained to be equal, and in which all a(i)'s were equal and all c(i)'s were equal. Comparison of the latter two restricted models with the full model of equation 3 allowed tests to be made of the degree to which the observed scales in different specialties differed from each other. The proportion of variance accounted for (R2) by the full equation 3 model was .858. R2-values for the all a(i)'s equal model was .835 and for the all a(i)'s equal and all c(i)'s equal model was .794. All of these R2-values were significantly greater than zero. The all a(i)'s equal R2 was significantly greater than the all a(i)'s equal and all c(i)'s equal model ($F(7,207) = 7.25$, $p<.05$). Furthermore, the full model R2 was significantly greater than that of the all a(i)'s equal model ($f(7,200) = 4.78$, $p<.05$). In sum, differences among specialties in scale use accounted for small but statistically significant proportions of the overall strength rating variation.

Table 2 presents values of the a(i)'s and c(i)'s estimated for the various specialties.

Table 2

Specialty Parameter Values

Specialty	a(i)	c(i)
112X0	3.061	.913
114X0	3.709	1.376
115X0	1.222	5.468
316X2F	4.564	.176
472X2	2.751	2.471
545X0	1.273	4.312
551X0	3.409	1.587
811X0A/X2A	2.949	1.936

Overall, the method proposed here for benchmarking task factor data appears to be feasible. Numerical optimization techniques were able to obtain least-squares estimates of the model parameters in the present example. The overall fit of the model was extremely good, and the scale use difference parameters, the a(i)'s and c(i)'s, accounted for a small, but statistically significant proportion of the overall rating variance. Further research is needed to investigate application of the benchmarking procedure to additional situations, in order to further explore the usefulness of the method.

Reference Note

1. Gott, S. P. Synopsis of the assessment of physical job requirements in the Air Force, Part I: Policy specification of composite and summary demand indices. Brooks AFB, TX: U.S.A.F. Human Resouces Laboratory technical report, in preparation.

References

Burtch, L. D., Lipscomb, M. S., & Wissman, D. J. Aptitude requirements based on task difficulty: Methodology for evaluation (AFHRL-TR-81-34). Brooks Air Force Base, Texas: Air Force Human Resources Laboratory, January, 1982.

Duncan, G. T. An empirical study of Jackknife-constructed confidence regions in nonlinear regression. Technometrics, 1978, 20, 123-129.

Fox, T., Hinkley, D., & Larntz, K. Jackknifing in nonlinear regression. Technometrics, 1980, 22, 29-33.

Weeks, J. L. The development and application of measures of occupational learning difficulty. Proceedings of the 23rd Annual Conference of the Military Testing Association, 1981, 1250-1257.

A TASK LEVEL INVENTORY FOR DESCRIBING JOB READING

Robert Vineberg
John N. Joyner

Human Resources Research Organization

In recognition of the potential mismatch between the literacy skills of personnel entering the Armed Services and the literacy demands of their jobs and training, the Services have undertaken a variety of programs of literacy research and development. One goal of these efforts has been to define the level of reading skill required to perform satisfactorily in different occupational specialties. This has generally been done by relating some index of the reading demands inherent in job performance (or in training) to a measure of reading ability. Several procedures have been used. The most frequent has been to relate the structural characteristics of prose passages in job reading materials to reading comprehension scores. This can be termed the <u>readability approach</u>. In it reading demands are estimated by applying readability formulas to samples of publications found in jobs or training settings. Such formulas typically translate features such as sentence length or number of syllables per word into an index of difficulty such as reading grade level. The Air Force in particular has used this method (Burkett 1976). Efforts to measure and improve the comprehensibility of text can be viewed as an extension of this approach.

The readability approach does not investigate the relation between reading performance and job performance; it takes as given that incumbents should be able to comprehend the publications found in jobs. The reading requirement of a specialty is defined as the difficulty of comprehending these textual materials, and the primary factor held responsible for differences in reading difficulty is the structural nature of the material read.

An approach that does address the matter of a job performance criterion can be referred to as the <u>job proficiency approach</u>. In this method reading comprehension scores are related to some criterion of job performance or proficiency such as job knowledge or performance test scores or supervisor ratings of performance. In using this method Sticht et al. (1972), for example, defined the functional literacy requirement of the specialty as that reading grade level at which no more than one quarter of job incumbents were found to be among the lowest quartile of performers on job sample tests. This method faces the problem of defining and measuring an acceptable criterion. Also, depending on the criterion that is used, the approach can be rather costly and may not be appropriate for general use. Undoubtedly the major difficulty with this method is the need to establish a causal link between reading ability and job proficiency. Unless reading is observed to be an actual part of job behavior, one hesitates to conclude that a relationship between reading ability and job proficiency is not due to some third factor such as general intelligence.

An alternative to both the readability approach and the job proficiency approach is the <u>job reading task approach</u>. Here reading tasks in a job are

first identifed and classified. Then tests incorporating a sample of these tasks are constructed to assess job reading skill.

Both the readability and the job reading task approaches, then, define literacy demand ultimately in terms of the comprehension of job reading materials. The latter approach is more refined, since the comprehension tests consist of tasks more nearly like job behavior than those on standard comprehension tests and since the measurement of comprehension is more direct, in that the intermediary of a readability formula is not used.

The job reading task approach to defining functional literacy is subject to its share of limitations. One problem is the metric for equating functional literacy levels across occupational specialties. In order to generate an index for this purpose, performance on the job reading task tests for particular occupational specialties has been related to scores on standardized reading tests (Sticht et al., 1971; Caylor, Sticht, Fox, and Ford, 1973). The literacy demand of a specialty is then defined as the reading grade level associated with any given criterion level of performance on the job reading task test. For example, Sticht et al. found that, if 80% of incumbents were expected to score at least 70% on the job reading task test, then the functional literacy level of an Army cook's job would fall between reading grade levels 7.0-7.9. So, the eventual expression of functional literacy requirements is less direct than the original specification of job reading tasks; the methodology does not entirely escape dependence on reference to general reading skill.

A second limitation of this approach is that it does not yield information about the relative difficulty of different job reading tasks themselves, nor about the relative representation of different tasks among different occupations. If one occupational specialty is found to have a lower literacy skill requirement than another, as indexed in terms of general reading grade level, it is not known whether this is due to a lower proportion of more difficult reading tasks or to some other factor affecting reading difficulty. Thus, although the approach starts by identifying job reading tasks in several specialties, differences between specialties are not made explicit, but are captured and represented only implicitly in the various job reading task tests.

Finally, the job reading task approach is a research method. Conducting interviews and developing a job reading task test for each specialty under investigation are costly. For this reason, the methodology is not suitable for application across a wide range of specialties.

In all of the methods I have just reviewed an objective measure of an incumbent's reading ability is introduced at some point. A procedure that provides direct, subjective estimates is the inventory approach. Sticht et al (1976), for example, used an inventory to identify the frequency of two particular classes of tasks in the Navy: fact finding and following directions. To estimate the difficulty of these tasks he returned to the reading task test approach and found evidence that reading to "follow directions" was more difficult than reading to "find facts."

The work I would like to describe this morning is a pure inventory approach to the identification of job reading demands. In it we obtained

subjective ratings of reading difficulty in addition to information about purposes for reading, criticality of reading, and the types of materials read. Our approach can be seen as analagous to obtaining ratings of difficulty of performing job tasks, as is done in occupational analysis.

We were seeking a method for estimating reading demands that would be compatible with Air Force occupational survey methods and that could be implemented readily without placing significantly greater demands on existing resources.

Evidence of the usefulness and dependability of the inventory would be sought in its capacity to detect differences in reading requirements across Air Force career ladders, in the extent of agreement among incumbents in an occupation about their reading, and in the extent to which the kinds of reading reported conform to expectations based on the nature of the tasks performed. As an obvious example, clerical personnel would be expected to report reading to transcribe, or type, more frequently than would aircraft mechanics.

In pilot work in a variety of different Air Force occupational ladders we found a task-specific reading inventory was capable of being readily constructed and administered. In such an inventory, information about job reading is obtained with regard to each one of a set of individual tasks that appear in conventional occupational analysis inventories.

Forms of the reading inventory were developed for job incumbents in the Airlift/Bombardment Aircraft Maintenance Career Ladder (AFSC 431X2) and the Administration Career Ladder (AFSC 702X0). In order to maximize the number of instances of reading in a field trial, the inventory in each career ladder called for information about the 40 tasks performed by the largest number of incumbents. Five questions are asked about each task. First, "The last time you did this task, did you do any reading?" Response options list seven purposes for reading. Second, "If you did any reading, how difficult was it?" A seven-point scale is used to rate difficulty. Third, "Did you need any help to understand what you read?" Fourth, "If instructions are needed for doing this task, can they be obtained without reading?" Fifth, "If you do any reading in this task, what materials do you read?" Response options list 13 types of reading materials.

Field Test Results and Discussion

In a trial, the inventories were administered to 169 incumbents in the maintenance ladder and 257 incumbents in the administration ladder. Data from the trial are based on approximately 4,500 occurrences of incumbent/task performance in the maintenance ladder and approximately 6,000 in the administration ladder.

Frequency and Purpose

In both occupations there are differences between tasks in percentage of incumbents who reported reading for task performance, ranging from 95% to 17% in AFSC 431X2 and from 95% to 22% in AFSC 702X0. If these differences among tasks prove to be stable, the inventory could be of considerable practical value in identifying aspects of job performance where reading is especially

important. In the present study, the sample was not split to permit estimating the consistency of the findings. Repeat administration of the inventory to additional samples is warranted.

In many cases, it is reasonable to infer the nature of task content from the task title, and the magnitude of reported reading generally appears to be appropriate to the nature of the task. In AFSC 431X2, for example, the two tasks with the largest percentage of persons reading are "Locate part numbers in illustrated breakdowns" (95%); and "Defuel aircraft using single-point method" (92%), where safety requirements prescribe that defueling be done in accordance with a written checklist.

Responses do not conform completely, however, to the expectations of such rational analysis. For example, only 75% of performers of the task "Edit drafts of administrative communications" reported reading, even though editing implies reading. Although occasional error of this magnitude seems tolerable, it may also indicate that response options for additional reading purposes should be added to the inventory. Reading to edit had been listed as a purpose in earlier trial versions of the inventory but was omitted from the final version to simplify it. A person who reads for the purpose of editing but who finds no such option on the inventory might mark "No reading done." While it is not possible to list all possible purposes for reading in the inventory, some changes in the current options may be desirable.

Other evidence of the dependability of the data obtained with the inventory is found by comparing the purposes given for reading to expectations based on the nature of the occupational tasks. In the maintenance ladder, for example, the purpose of reading to look up facts is given by 81% of persons performing the task "Locate part numbers in illustrated parts breakdowns," a purpose clearly implied by the task title. The percentage of persons who indicate reading in this task for other, less obvious reasons range from only 8% to 36%.

Some type of reading was reported in 61% of the occurrences of incumbent/task performance in the maintenance specialty and in 67% of the occurrences in the administration specialty. From 24% to 67% of job-related reading is done to look up facts, find out that a task is to be done, or to learn or check the procedure for carrying out a task.

Incumbents in the maintenance specialty showed greater agreement about their purpose for reading than those in the administration specialty. In the maintenance ladder, 73% of the reliability coefficients computed for estimating agreement about reading purpose were .9 or above whereas in administration 32% were .9 or above and 69% were .8 or above.

Type of Material

In the maintenance ladder, 56% of task performance occurrences include reading work cards, job guides, and inspection cards. In the administration ladder, these same materials are consulted only 11% of the time. Publications such as manuals, technical orders, and regulations provide the most frequent reading content in this specialty.

Major differences in the materials in which reading occurs in the two specialties are found in the categories of work cards, job guides, and inspection cards (56% in AFSC 431X2; 11% in AFSC 702X0), messages, letters, TWXs, TCTOs (12% in 431X2; 21% in 702X0), and in material to be copied, typed, or reproduced (6% in 431X2; 22% in 702X0). Like purposes for reading, these differences in types of material read conform to generally expected differences in the work requirements of the specialties.

As with purpose for reading, when the types of material read are examined by individual task, they appear to be appropriate to task content. For example, in maintenance, 90% of those who had performed the task "Inspect landing gear components" reported using work cards, job guides, and inspection cards.

In summary, incumbents appear to discriminate as well or better between types of material read as between their purposes for reading.

Reading Difficulty

The ratings of difficulty of reading in individual tasks proved to be the aspect of the inventory of least certain usefulness. While there were differences across tasks in both specialties, agreement among incumbents about the level of difficulty was poor. The reliability of the average reading difficulty rating for tasks in the maintenance specialty was .24 and in the administration specialty, .74. Mean difficulty ratings in the two specialties were not significantly different though perhaps they should have been. It can be estimated on the basis of their higher AFQT scores that incumbents in the maintenance ladder possess higher reading comprehension scores than those in the administration ladder. Ratings of reading difficulty, therefore, might be expected to reflect this difference. On the other hand, AFQT also decreases with grade and increasing difficulty of reading is reported with increasing grade in both specialties. Since the readability level of the printed material read by respondents in each specialty is unknown, we can only speculate as to whether the lack of difference in perceived difficulty of reading between specialties is due to the observed unreliability of the ratings, a compensating difference in the difficulty of reading materials, or some other factor.

Need for Assistance

As I mentioned earlier, the inventory also included a more concrete measure of job reading difficulty: the need for assistance in reading.

In contrast to the ratings of reading difficulty which showed little difference between specialties, there is a greater need for assistance in reading in the administrative specialty. This finding is consistent with the estimated lower reading ability of incumbents in this occupation. Although incumbents in this specialty did not rate their reading as very difficult, they reported needing help to understand it an average of one out of every eight times they read something in task performance. By contrast, those in the maintenance specialty reported needing help in reading less than one time in twenty.

CONCLUSIONS

The principal objective of this research was to develop an inventory approach to estimating Air Force job reading requirements. We have concluded that the approach is feasible for further development and implementation. The inventory was readily compiled from existing occupational analysis data, mass produced, mass administered using current Air Force procedures, optically scanned, and analyzed using existing Air Force equipment and resources. It is effective in capturing differences in reading requirements and behavior between specialties and among job tasks within a specialty. On the basis of job task titles and some obvious differences between maintenance and clerical occupations, the kinds of reading reported appear appropriate to the nature of the incumbents' activities. The fact that incumbents do discriminate reading requirements across tasks indicates that data from the reading inventory could be used conjointly with data collected in the Air Force Occupational Analysis Program. Apart from a variety of modifications and adjustments to the inventory the most compelling requirement is a need to administer the inventory to additional samples of incumbents to determine the stability of the findings and thereby the dependability and value of the instrument for operational use.

Work Cited

Burkett, J. R. A review of research and development on literacy and technical writing in the Air Force. In T. G. Sticht & D. W. Zapf (Eds.), Reading and Readability in the Armed Services, HumRRO FR-WD-CA-76-4. Alexandria, VA: Human Resources Research Organization, September 1976.

Caylor, J. S., Sticht, T. G., Fox, L., & Ford, J. Methodologies for determining requirements of military occupational specialties. HumRRO TR-73-5. Alexandria, VA: Human Resources Research Organization. March 1973.

Sticht, T. G., Caylor, J. S., Kern, R. P., & Fox, L. C. Determination of literacy skill requirements in four military occupational specialties. HumRRO TR-72-23. Alexandria, VA: Human Resources Research Organization, November 1971.

Sticht, T. G., Fox, L. C., Hauke, R. N., & Zapf, D. W. Reading in the Navy HumRRO-FR-WD-CA-76-14. Alexandria, VA: Human Resources Research Organization, September 1976.

AD P000915

How To Display Data Badly

Howard Wainer
Educational Testing Service
Princeton, New Jersey 08541

Summary

Methods for displaying data badly have been under development for many years, and a wide variety of interesting and inventive schemes have emerged. Presented here is a synthesis yielding the twelve most powerful techniques that seem to underlie many of the realizations found in the literature. These twelve (the dirty dozen) are identified and illustrated.

This is the text of an address to the Military Testing Association (MTA) at its 24th Annual Conference in November 1982. It was supported by the Program Statistics Research Project of the Educational Testing Service, and a full copy of the manuscript can be obtained by requesting from the author the ETS Program Statistics Research Technical Report RR No. 82-35.

RULE 1: Show as little data as possible (minimize the date density)

Figure 1

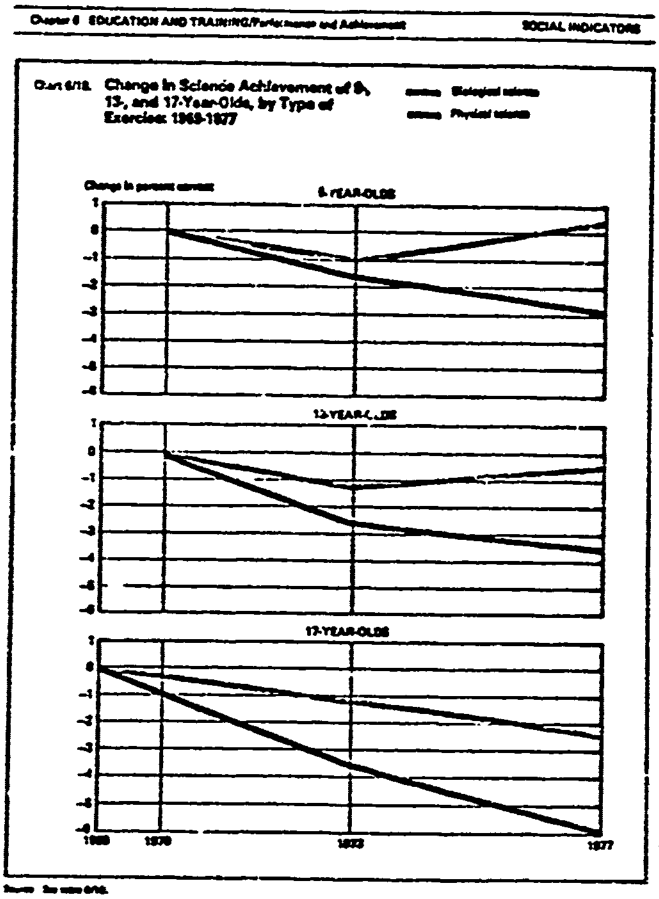


Table 1

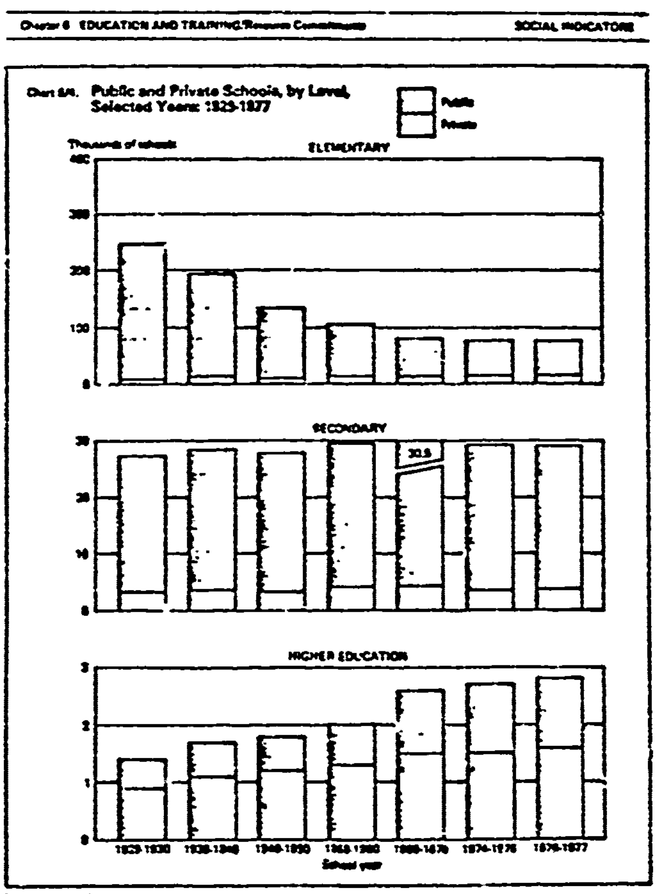
Data Density in Twenty Randomly Selected Statistical Graphics for Some Recent Publications (in number of numbers per square inch)

	Median	Minimum	Maxim[um]
Nature	48.0	3.0	362.0
Journal of the Royal Statistical Society, B	27.3	4.1	178.5
Science	20.8	4.6	46.0
Wall Street Journal	19.2	3.1	136.4
Fortune	17.9	5.3	31.2
The Times (London)	17.8	1.9	122.3
JASA	16.9	4.0	168.5
Le Monde	8.1	0.5	16.6
Psychological Bulletin	7.5	0.8	73.7
Journal of the American Medical Association	7.4	0.9	38.7
The New York Times	7.4	1.0	13.2
Newsweek	5.8	0.5	15.0
Scientific American	5.1	0.5	69.4
Statistical Abstract of the United States, 1979	4.6	2.0	23.0
Social Indicators III	0.6	0.2	1.8
Pravda	0.2	0.1	1.0

Source: Tufte, 1982

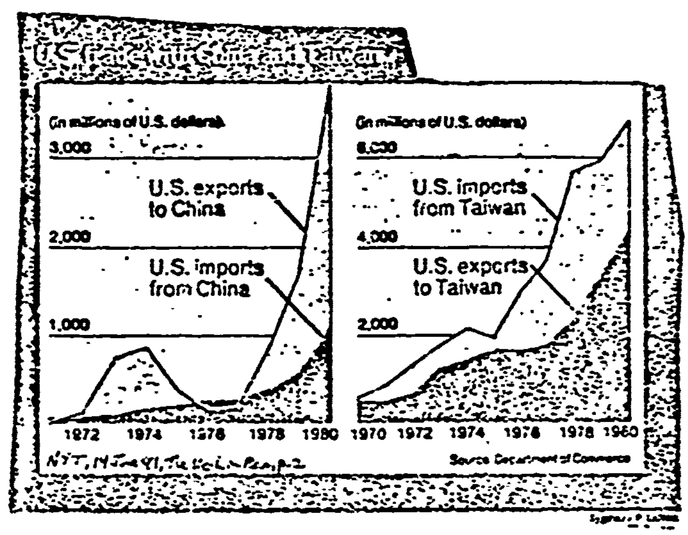
RULE 2: Hide what data you do show

Figure 2



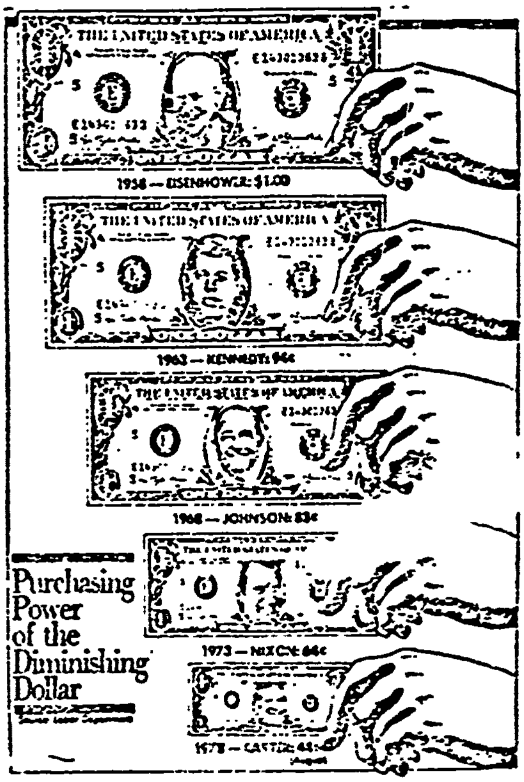
RULE 3: Ignore the visual metaphor altogeth[er]

Figure 3



RULE 4: Only order matters (The Pravda School of Ordinal Graphics)

Figure 4

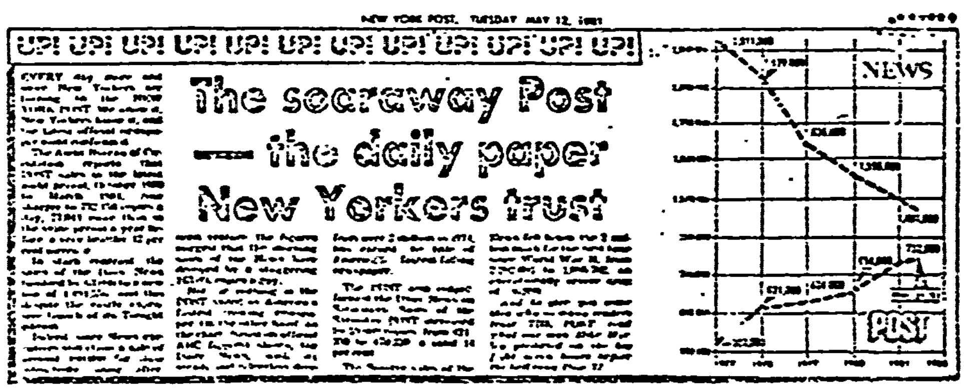


RULE 5: Graph data out of context

Figure 5

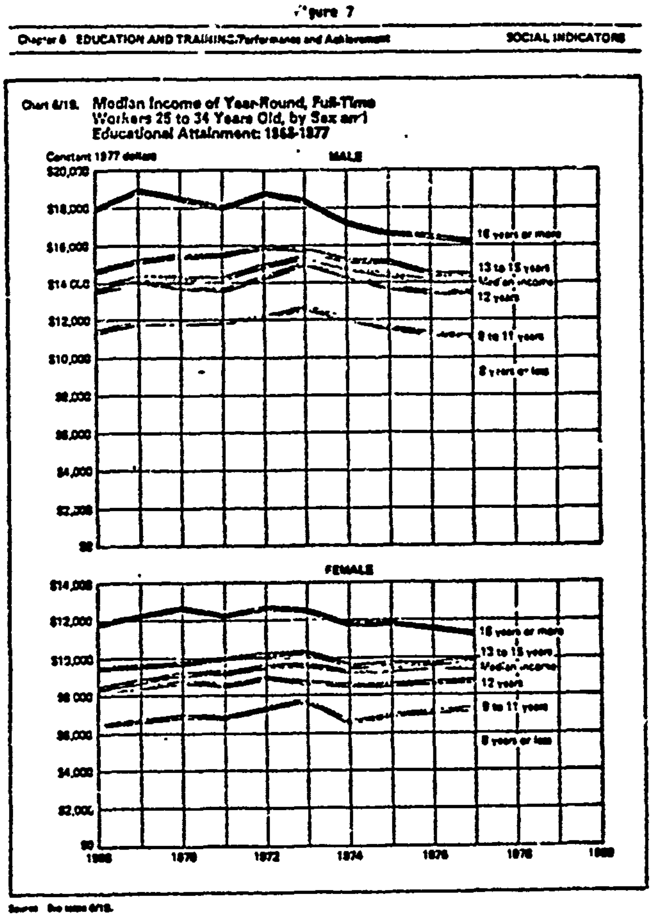
RULE 6: Change scales in mid-axis

Figure 6



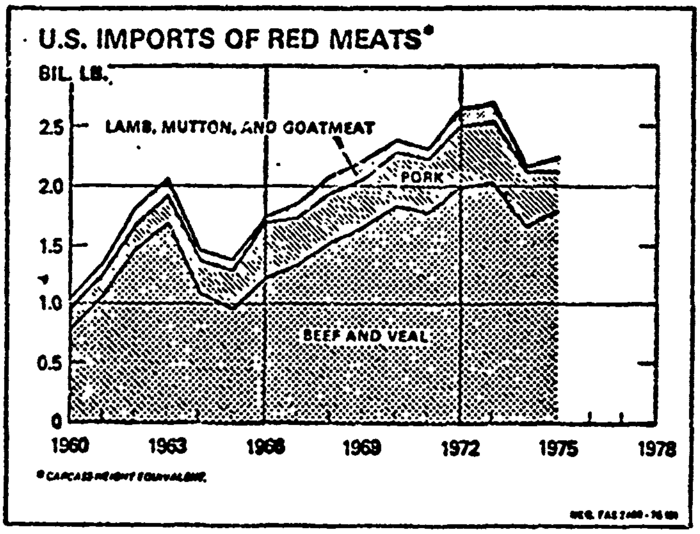
RULE 7: Emphasize the trivial (ignore the important)

Figure 7



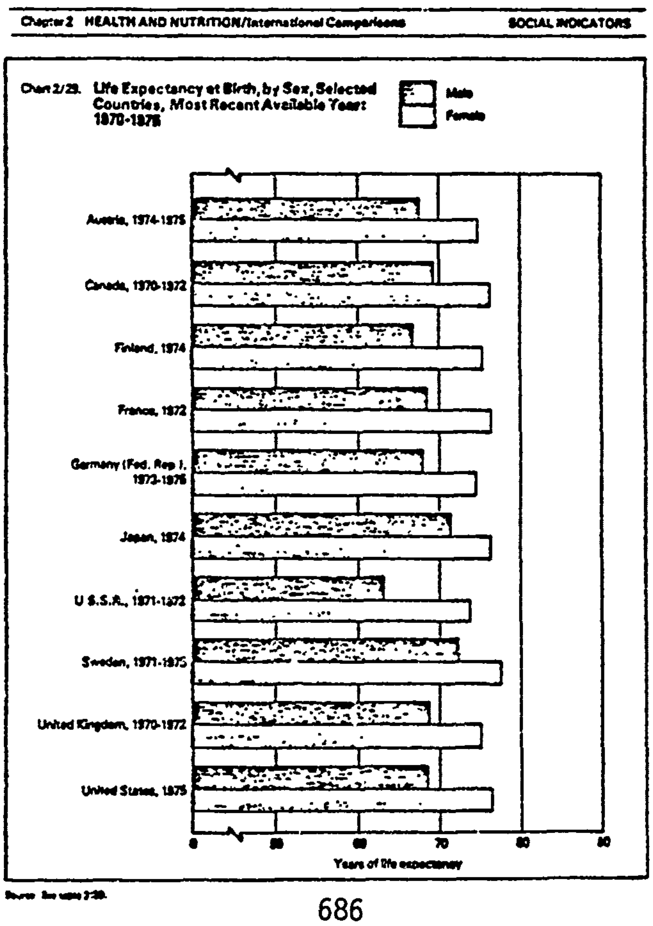
RULE 8: Jiggle the baseline

Figure 8



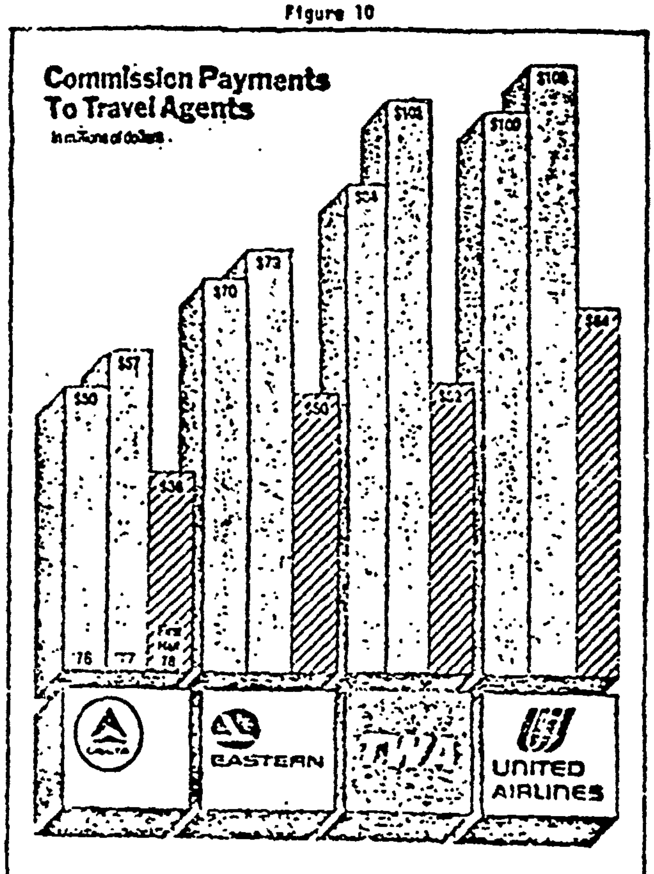
RULE 9: Alabama First!

Figure 9



RULE 10: Label: a) illegibly, b) incompletely, c) incorrectly, and d) ambiguously

Figure 10



Complex web of discount fares and airlines' telephone delays are raising travel agents' overhead, offsetting revenue gains from higher volume.

RULE 11: More is murkier: a) more decimal places; b) more dimensions

TABLE 2

Optimal Selection from a Finite Sequence with Sampling Cost

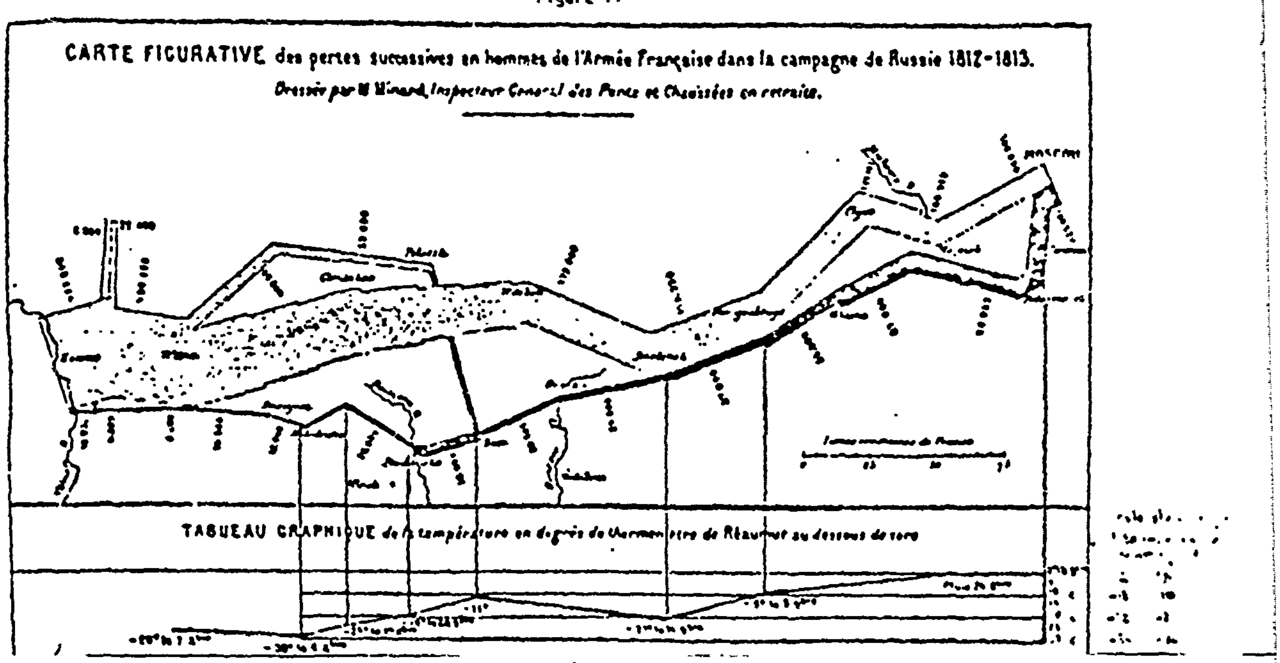
$$g(X_{s+r-1}) = \begin{cases} bE(X_{s+r-1}) + a, & \text{if } S = s \\ 0, & \text{otherwise} \end{cases}$$

	b/c = 10.0		100.0		1000.0	
N	r*	$(C_N(r^*) - a)/c$	r*	$(C_N(r^*) - a)/c$	r*	$(C_N(r^*) -$
3	2	.20000	2	2.22500	2	22.47499
4	2	.26333	2	2.88833	2	29.13832
5	2	.32333	3	3.54167	3	35.79166
6	3	.38267	3	4.23767	3	42.78764
7	3	.44600	3	4.90100	3	49.45097
8	3	.50743	4	5.57650	4	56.33005
9	3	.56743	4	6.26025	4	63.20129
10	4	.62968	4	6.92358	4	69.86462

Source: Dhariyal & Didewicz, 1981

RULE 12: If it has been done well in the past, think of a new way to do it

Figure 11



References

Tufte, E. R. The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press, 1982.

Wainer, H. Making newspaper graphs fit to print. In Bouma, Kolers & Wrolsted (Eds.), Processing of Visible Language, (Vol. 2). New York: Plenum, 1980.

Wainer, H. How are we doing? A review of Social Indicators III. Journal of the American Statistical Association, in press.

Wainer, H., & Thissen D. Graphical data analysis. Annual Review of Psychology, 1981, 32: 191-241.

AD P000916

A PERSONNEL TEST OF FIREFINDER RADAR TRAINING, TRAINER, AND SYSTEM

Raymond O. Waldkoetter and Richard B. Modjeski
U.S. Army Research Institute for the Behavioral and Social Sciences
Alexandria, Virginia 22333

Analysis of earlier automated training development needs appeared constrained. Certain training design limitations were imposed by a rapid system development schedule for the Field Artillery's AN/TPQ-37 Firefinder Radar and its A17E11 training device. A training device was proposed to reduce costs that would accrue if soldiers were trained only on an operational system, where numbers trained would be limited and length of time to train hard to control. Such a training device can concentrate attention and effort to accomplish a better integrated instructional process with attainment of individual and crew task objectives.

With the accelerated acquisition program, early training design documents could not fully accompany the radar and device. Development of test and evaluation acceptance programs was seriously curtailed. These programs could furnish training design guidelines if planned with greater detail in human factors and personnel support for the actual equipment and device design products. An inventory form for Firefinder training requirements was constructed to alleviate part of the information constraint. The form was oriented on the system training device to evaluate the utility of the commitment to simulated training and transfer to the AN/TPQ-37 Radar. This form was also the primary instrument for a personnel test of training and the Firefinder systems by verifying training design specifications and prior results from concept evaluation and user test phases.

A convergence of Firefinder course design and test requirements occurred at the earlier concept evaluation and user test phases (Lovell et al., 1980) indicating two apparent conditional training constraints. First, there were limited data available about what should constitute testable training on the A17E11 device and AN/TPQ-37 Radar. Secondly, the user test was compelled to base evaluation of training and system suitability on part of the first conditional constraint findings, while extracting training measures of effectiveness subject to continuing revisions in content and performance standards. Under these conditions there were training issues and measures that were not tested at a desired level of precision. Student, instructor, training device and equipment system relationships were not identified sufficiently to evaluate which tasks, operations or system features defined the best test of training system capability.

Other information-gathering alternatives were not proposed due to the pressure of pending training development and system test schedules. Neither an analysis approach nor model could concurrently evolve which would more economically "test" a small number of operators or mechanics in a manner similar to a

---

The views expressed in this paper are those of the authors and do not necessarily reflect the view of the U.S. Army Research Institute or the Department of the Army.

structured "test pilot" evaluation (Kratochwill, 1978). Naturally, design of performance criteria should begin with the system creation. Later evaluation of personnel test requirements and training can proceed directly from those documented design guidelines which specify human factors and personnel requirements for system engineering and functional operations. One proposed acquisition, test and evaluation, and training development system has already been demonstrated. It would design each performance requirement with simulator specifications and man-machine interface controls (Hritz & Purifoy, Jr., 1980).

Thus a few carefully selected test-players could reliably exercise a system's operational capabilities and automated training requirements in a completely instrumented scenario, when guidelines specify human factors and personnel requirements. Group test evaluation procedures are now relatively anachronistic if measuring only operational task behaviors. A training device such as the A17E11 can test a limited number of personnel for both learning and operational tasks. But it must also display a high degree of fidelity, satisfy rigorous design parameters, and have full performance evaluation guidelines for personnel test and training procedures.

Personnel training effectiveness of the training program for the A17E11 Firefinder radar trainer and AN/TPQ equipment was evaluated by an interview-survey form developed to examine training policy needs. This 81 item form was given to 53 personnel selected as test-player subjects. To augment information limited by the accelerated systems acquisition, the form was analyzed as a system "personnel test" by group, background, and question (item) variables for learning task effects on trainer A17E11 or AN/TPQ-37 Radar training. Test subject responses were used to suggest training policy revisions using expected operator tasks and observed deficiencies.

METHOD

A questionnaire approach to analyze training development needs and personnel consequences is not unique. A comprehensive review format, however, was newly formulated to recover performance objectives rather implicitly expected in the systems' design. There is an innovative procedure, additionally, in gathering and synthesizing information for course design which was not previously referenced nor based on immediately observed training conditions. Moreover, the methodology application has pointed to finding a further clarification and coherent integration of system design procedures. Such procedures should project specified training guidelines and personnel test requirements so that an economical and accurate strategy will guide the parallel activities of Artillery system acquisition, test evaluation, and training development.

If total coordination of system design, test, and training task objectives is conceptualized and implemented, simulator and equipment systems should fully demonstrate any designed operational features. Any suggested modification data are, then, still acceptable as system test, personnel and training decisions are formulated well before system installation. To support this adaptive concept requires, also, the early selection of a centralized coordinator to direct and monitor every critical aspect of acquisition, test and training requirements to deliver effective decisions for system design and training. A coordinator must possess the stated responsibility to intercede anytime to effect the required decision processing of either institutional managers, technical experts or contractual support personnel.

An alternative analysis approach was prompted by events to bring a degree of synthesis between course design objectives and test sanctions. This would better accommodate training and clarify test results for any course, device, or equipment changes. Questionnaire acceptance suggests that the personnel test instrument was effectively constructed to describe student concerns and equipment system relationships. These features were noted from technical observations later verified by the Firefinder training device item responses given by students and instructors and a critique of technical reviewers. Though a "one-time" instrument, the design review format for training analysis may suggest some type of standardized approach by which critical training issue and equipment capability measures can move toward increased utility and precision. Other alternative analysis approaches will surely evolve for training device acquisition and training development as more advanced computerized training systems are requested.

As an example of possible generic dimensions following from the questionnaire structure, some standard content features examining the A17E11/AN/TPQ-37 systems were projected. These generic dimensions developed on the given systems were then applied tentatively in an evaluation of the A17E14 Firefinder Maintenance Trainer. Where write-in or interview comments appeared very briefly because of the highly detailed survey analysis, there is no suggestion to pursue collection of observer or test player remarks during test operations, except as an analyst may wish to annotate some condition.

An interim analysis instrument as the method advocated in this report could yield significant training design information for review of course content and simulated performance criteria. When training information documents may have omitted certain simulated and prime system training and instructional guidelines during accelerated development, an auxiliary effort is justified. That effort should construct a training inventory and interview form to obtain the best personnel test data available. Developing a flexible questionnaire format to interpret user/operator transactions with an automated system (Berger & Hawkins, 1979), furnishes a viable alternative to support ISD system acquisition and training development activities under constrained conditions. This approach is illustrated in that simulated training device/equipment operations and personnel training needs were effectively augmented for the Firefinder Radar Systems.

A progressive review of system acquisition and training design testing would be conducted by applying the instrument results using interface perceptions of instructor, student and training device/equipment. Evolving content and form design questions which arose intimated a possible instrument combining eventually personnel test and task inventory capabilities.

Research questions explored completion of performance objectives, proficient trainer transfer to the AN/TPQ-37, tasks trained and deficiencies. These questions were designed to generally answer whether this personnel test of the implemented training system could better integrate trainer (A17E11) performance in the actual MOS 13R10 course. Group responses could indicate significant preferences for training policy activities, course content, proficiency needs, and augment already proven systems. Minimal background variables might affect responses on training performance standards while suggesting remedial training tasks. Though operator skills may be perceived as difficult to learn, tasks were to be identified for a revised task sequence and correct operational procedures to achieve proficient skill within critical learning times.

RESULTS AND DISCUSSION

Research questions stated for the personnel testing of the training effectiveness and transfer in the Firefinder course, resulted in a generally positive set of findings for training design activities and A17E11/AN/TPQ-37 operations. These questions were intended as goals by which to analyze progressive achievement in training development. They have suggested modifications to support continuing training competence and course improvements. Questionnaire evidence and intensive two-year observations by the researchers tentatively found that probable training effectiveness for student operators could be fully expected. Certain course insights and implied modifications can work to furnish an optimal training program. Improved training development and device requirements were being defined during the implementation phase of trainer acquisition and instruction. Questionnaire items were analyzed by percentages and the chi square test of significance (.05 level) examined for each of the item cross tabulations with associated correlations.

Performance Objectives. Questionnaire items affecting performance objectives were analyzed noting whether these items would describe learning constraints or options to choose an effective training-task solution. Operators are expected, it appeared, to attain or exceed performance objectives for the A17E11/AN/TPQ-37 systems when complementary tasks are explained, course content is made pertinent, and instructor skills are evident. Items 67, 60, and 50 were interpreted as specifically conveying the confirmed findings for the first research question. Performance objectives in course achievement could then be further attained or exceeded as given tasks and operations were exercised in the proper sequence. Operators succeeded, responses indicated, as instructors displayed necessary skills and helped students on the trainer and equipment, referring to manuals and radar experience. Students learned faster and better utilized study time to complete performance objectives, responses agreed, when the training sequence applied the best mix of trainer/equipment practice and study materials. Item 38 reflected a relatively conclusive overview with 98% of the test subjects significantly acquiring "reasonably to very sufficient skill" on the trainer to operate the actual equipment. This finding additionally reinforces the cumulative transfer evidence given below.

Trainer Equipment Transfer. Proficient trainer performance was expected from responses to transfer to the AN/TPQ-37 Radar and result in successful operation. Test subjects answered item 30 by a significant majority (76%) agreeing to the performance similarity of the systems and procedures. Where personal background of the test students showed some significant differences, this majority observation was still upheld. A contrast on item 24 was shown by the effects of group background differences. Here differences were experienced by the test personnel in that their " 'make-up' study to reach required proficiency standards" reflected some individual training preferences and course flexibility. A compressed training schedule seemed to affect the responses to item 68 regarding whether training on the actual system was more effective and useful than on the trainer. Responses tended to favor training on the actual equipment which may simply capture the preference of the test players preparing for their test site. Also the interesting conclusion is implied that test students had enough short-term training experience to compare system experiences and then prefer AN/TPQ-37 training over initial A17E11 training. The group response to item 57 showed about 74% expecting to need AN/TPQ-37 proficiency training "monthly" or more often. Researcher observations were used to analyze this relationship suggesting subjects were significantly aware of A17E11/AN/TPQ-37 transfer skills needing practice in the unit location

to complement resident training. Transfer from the A17E11 to AN/TPQ-37 was facilitated by proficient map-reading and radar skills, it was noted, and may be most handicapped if a student has low reading skill and below average mental ability (item 81).

Tasks Trained and Deficiencies. A narrative for training performance standards described in terms of items, what was trained effectively and deficiencies needing further training development according to nine content factors. Review of item responses permitted an evaluation that the course development process had succeeded in designing critical performance sequences. Guidance furnished from this process was used to adjust proficiency standards in reference to prior device acquisition and development requirements. Training of critical tasks and identifying deficiencies were predicated on relating other items (43, 46, and 54) for example, using group difference and background variable difference. The content factor results gave a unifying perspective, while a research question analysis probed other item relationships affecting control of training performance effects. If instructors explained task differences and assured availability of training materials and feedback evaluation of student errors with increasing efficiency, answers agreed, a firm basis was prepared to control critical task learning and correct deficiences. In spite of some background variable differences for item 32, a significant consensus was still obtained to report complete enough "field training to learn the required operational tasks for the AN/TPQ-37." Certain deficiencies were experienced relating to time in the primary MOS, time in the Army, and rank. Item 34 gave an overview evaluation for A17E11 task training and guidance with nearly 100% of the test students answering that "usually to completely adequate" monitoring of student errors was given to direct feedback and correction.

Trainer-Course Testing. Training effectiveness testing (Finley & Strasel, 1978) of the Firefinder course increased the understanding of trainer features and learning tasks, respondents agreed, to better integrate it in the course delivery and with the AN/TPQ-37 Radar system. The related research question was confirmed by a number of associated findings. Training requirements for the A17E11 were studied more (item 47), replies agreed, as the course was improved by on-going training design changes. It was conceded (item 73) that the instructor-student ratio of 1 to 6 should approach 1 to 3 to increase the attention level and interest. That some instructor-console tasks could require most of the instructor's time (item 69) was largely rejected by test-player answers, but less so the longer away from school radar training. The instructor-console function needed further development, it appeared, to maximize student A17E11 simulation activities. Generally test subjects significantly observed the effective A17E11 "course sequencing" with about 75% replying they were able to make suggestions improving A17E11 instruction (item 26) and instructors were more often able to answer A17E11 questions to maintain the training progress (item 40). The positive evidence presented for the other research questions above was also accepted as reasonable support for a positive answer to this last question area.

In summary findings indicated modifications acceptable to continuing training program development for the A17E11/AN/TPQ-37 systems. The personnel test more clearly described how the training program could maximize the already engineered potential for trainer/radar training effectiveness and transfer features. Training policy decisions were derived from research observations sampling performance standards. Support was provided for an improved training design and device acquisition process at generic and system specific levels.

REFERENCES

Berger, B. M., & Hawkins, H. H. Occupational analysis: An automated approach. In T. Abramson, C. K. Tittle, and L. Cohen (Eds.), Handbook of vocational education evaluation. Beverly Hills, CA: Sage Publications, Inc., 1979.

Finley, D. L., & Strasel, H. C. "Training Effectiveness Testing (TET) in training device development." Paper presented at the USDRE Training and Personnel Technology Conference, Washington, D.C., 11 May 1978.

Hritz, R. J., & Purifoy, G. R., Jr. Maintenance training simulator design and acquisition (AFHRL-TR-80-23). Lowry Air Force Base, CO: Air Force Human Resources Laboratory, August 1980.

Kratochwill, T. R. (Ed.). Single subject research: Strategies for evaluating change. New York: Academic Press, 1978.

Lovell, J. A., Klinger, D. R., Trovato, A. E., McLeod, E. S., Raney, R. W., & Meisenzahl, K. A. On-site user test of Firefinder radar operator/maintenance trainers, device A17E11 and device A17E12 (Final Report ACN 24488/80 OTN 435). Fort Sill, OK: U.S. Army Field Artillery Board, July 1980.

AD P000917

THE NROTC INSTRUCTORS SEMINAR: PREPARING OFFICERS FOR COLLEGE TEACHING

Commander John B. Washbush, USNR, Executive Officer NROTC Unit and Associate Professor of Naval Science, University of Wisconsin, Madison, WI (previously head of professional development for the NROTC, Headquarters, Naval Education and Training Command)

BACKGROUND

For the past 37 years, the Navy has provided teacher training for naval officers being ordered to instructor billets at the Naval Reserve Officers Training Corps (NROTC) units. These Navy officers are usually lieutenants (0-3) rotating from their initial sea duty tours with operational units. Because of the mobility inherent in sea duty, young officers do not normally have the opportunity to complete post-graduate degree work while operational; however they do complete basic warfare specialty qualifications. In addition to Navy officers, Marine captains and majors are assigned to NROTC units as Marine Officer Instructors (MOI). The operationally focused backgrounds of these Marine officers are similar in nature to those of the Navy lieutenants.

At the educational institution the NROTC instructor is given faculty status in a department of naval science and teaches accredited courses. Thus, the officer needs appropriate academic training to round out operational expertise in preparation for the teaching role. The Navy has responded to this need by providing the officer with a special training program, the NROTC Instructors' Seminar.

SEMINAR ELEMENTS AND STRUCTURE

The Seminar is more than a "how to teach" program. It recognizes the reality of the multi-dimensional responsibilities of the NROTC instructor as teacher, counselor, academic advisor, program administrator, and role-model. The Seminar is structured into an intense two-week package which addresses each of these facets and attempts to build or enhance appropriate skills.

In its present state of development, Seminar contains these distinct elements:

a. Curricular Education: Instruction in the form and content of the courses the instructor will teach, including instructional resources.

b. Teaching Methods Instruction: An overview of instructional methods (including lecture, discussion, seminar, and teaching interview), evaluation and testing, psychology of learning, and the philosophy of teaching.

c. Instructor Competency Training: Identification and practice of

behavioral competencies which are believed to promote superior performance in NROTC instructors.

d. Training in NROTC Program Administration: A survey of student administration, program procedures, data systems, academic and training requirements, and unit-headquarters relationships.

e. Counseling: Instruction and intensive practice in interviewing methods aimed at helping students identify and resolve problems and make personal decisions; overview of typical college student problems; ethical issues in counseling.

f. Supervised Practice Teaching: Development and practice of applied instructional skills using NROTC curricular material. This is the integrative, capstone exercise for students.

As it is evolving, Seminar is becoming increasingly student-active. The focus is the development of skills which will promote the acquisition of performance abilities in order to benefit students and render impotent any arguments that NROTC instructors lack sufficient academic credentials.

Staffing Seminar requires using NROTC-program personnel. Typically, the officer-in-charge and the professional director are headquarters staff members. Additional administrative personnel and all course content-area instructors are supplied by field units. The content-area instructors are highly motivated and skilled officers who not only have demonstrated superb instructional skills but who have been headquarters chosen to act as points of contact at designated course-coordinator units. In addition to these officers, instructors in teaching methods and counseling have been available in the person of two naval reserve officers who are academic professionals. They have provided a haven of continuity and superb impact on student officers for the past decade.

FORCES FOR CHANGE

Until 1981, Seminar development was gradual and somewhat random. Additionally, a substantial amount of student passivity occurred--too much tell-them-how-to-do rather than doing. Two elements have converged during the past two years to establish a purposeful course toward an improved product. The first of these was a deep-seated concern, expressed over several years, that practice teaching was not sufficiently effective in promoting and testing student instructional-skill development. The second event was the impact of the Navy's Leadership and Management Education and Training (LMET) program. This program requires some elaboration.

The LMET program is a broad-based commitment by the Navy to identify and teach skills intended to improve the character of officer and enlisted leadership. The program model is one developed by McBer and Company, a consulting firm, and widely used with its clients. In brief, the method used seeks to identify behavioral skills and abilities which are purported to distinguish superior performers from average/poor ones. The critical incident interview is used as the data-gathering tool. Interviews are examined and scored to identify basic themes, and comparisons are made of interviews of designated superiors with those of the less able in order to isolate

different behaviors. Categories of behaviors are identified with differentiate performance and causality is attributed. The "competencies" are then taught in training courses. The research basis for the Navy LMET program is not, in the mind of this reviewer, very elegant, and certain assumptions might well be challenged; however the Navy has made so great a commitment to LMET, that such arguments in this paper would take us off the track. Suffice it to say, the job of the NROTC instructor has been identified by the Chief of Naval Operations as one for which this training will be provided. Thus it must be accommodated in the NROTC Instructors' Seminar.

The contractor conducted data gather at the Naval Academy, at the Officer Candidate School, and at a sample of NROTC Units. Data, in the form of critical incident interviews, were gathered on two billets, instructor and company officer (the NROTC instructor billet is an amalgam of these two, containing tasks of each). From this survey, McBer identified 16 competencies which are stated to be attributes of superior performance as an NROTC instructor. These competencies are:

1. Demonstrates Student-Centered Diagnosis
2. Takes Initiative
3. Sets High Performance Standards
4. Focuses on Results
5. Assesses Self Accurately
6. Clearly Communicates Abstract Ideas
7. Demonstrates Enthusiasm About Teaching
8. Creates and Uses Imaginative Teaching Strategies
9. Prepares Students for the Fleet
10. Influences
11. Demonstrates confidence in Personal Authority
12. Gives Negative Feedback
13. Demonstrates Self-Control
14. Demands Personal Responsibility
15. Demonstrates Positive Expectations
16. Understands

The logic applied to the LMET training program for NROTC instructors is as follows: Competencies are described in such behaviorly specific ways that they can be taught in training courses; training and practice in the competencies will produce mastery of them: superior instructor performance will result - supposedly! The training cycle used involves these stages: Recognition; Understanding; Self-Assessment; Skill Practice; and Application. In the 1982 Seminar, competency instruction carried through the entire training cycle the three teaching competencies (Clearly communicates abstract ideas; Demonstrates enthusiasm about teaching; Creates and uses imaginative teaching strategies) and Influencing. Other competencies were considered to be addressed to at least the recognition level in other parts of Seminar, especially the teaching methods and counseling courses. The competency training course itself used a combination of contractor personnel and officer staff members.

In addition to the competency training course, a considerably improved practice teaching program was devised and used in Seminar 82. Ten hours of total Seminar time was allocated to supervised practice teaching

during week two. Each student was assigned to a practice teaching group and received about one individual hour of platform time. Students made three presentations: An impromptu brief ice-breaker without critique; a 10-minute supervisor-critiqued presentation on any topic; a 20-minute class-critiqued lesson on an NROTC course topic, including a lesson plan. Evaluation forms used reflected and reinforced concepts from both the competency training and teaching methods courses. When not on the platform, students acted as class members and evaluators. Thus, practice teaching became the integrative element in the overall instructional program of Seminar.

PROBLEMS AND RECOMMENDATIONS

Evaluation of Seminar 82 revealed growing pains. Many of the difficulties noted are attributable to forcing the LMET competency training into an established program while attempting to retain a two-week training-cycle format. A large number of respondents stated strong negative reactions towards certain aspects of the competency training. Comments ranged from "waste of time" to "overkill". Excessive repetition of terms was also cited. These negative reactions not only document the need for continuing work at integrating the competency training into the other effective aspects of Seminar, but they also reflect real flaws in the competency program.

The most obvious flaw is overemphasis on an excessive number of competencies and far too many accompanying behavioral indicators. A neater and more logical package is needed. All competencies address two fundamental aspects of the NROTC instructors job: Communication and Feedback. Such categorizing would permit the essential elements of Seminar to be integrated in a meaningful pattern as follows:

<u>Communication</u> (content, structure and delivery)

Program Administration (policies, regulations, procedures)
Course Content
Lesson Planning
Counseling/Advising
Classroom Methods
Practice Teaching

<u>Feedback</u>

Evaluation and Testing (student, course, program)
Program Administration (data systems and reports)
Counseling/Advising
Practice Teaching

In addition, the 16 competencies may be grouped within the categories:

<u>Communication</u>

(2) Takes Initiative
(3) Sets High Performance Standards

(6) Clearly Communicates Abstract Ideas
(7) Demonstrates Enthusiasm About Teaching
(8) Creates and Uses Imaginative Teaching Strategies
(9) Prepares Students for the Fleet
(10) Influences
(11) Demonstrates Confidence in Personal Authority
(13) Demonstrates Self-Control
(15) Demonstrates Positive Expectations
(16) Understands

Feedback

(1) Demonstrates Student-Centered Diagnosis
(4) Focuses on Results
(5) Assesses Self Accurately
(12) Gives Negative Feedback
(14) Demands Personal Responsibility

The "forest and trees" problem outlined here is a common difficulty in modern education methods, whether competency or objective-based. We educators love to over-define, to note too many discrete elements, and to over-structure at the expense of integration. Here is the classic problem of parts and the whole. Too often, implicitly or explicitly, we assume summation of parts is sufficient. The results are predictable! In the case of the NROTC Instructors' Seminar, these problems have been made worse by the requirements to force a new program into a well-established model.

In this somewhat critical impasse, considerable potential exists for clarifying and improving the preparation of naval officers for NROTC instructor duty. The following actions should be taken:

(1) Role Define the NROTC Instructor's Billet. The officer must be able to understand and perform the roles of teacher, advisor, counselor, administrator, and role-model.

(2) Objectivize. State, as simply as possible, training objectives relating to each role. Emphasize integration and mutuality of objectives.

(3) Map the Program. Define clearly the means by which objectives are to be met. Give careful attention to both content and process.

(4) Structure the Program. Define each major instructional area and fix responsibilities for design and delivery. Highlight carefully-defined integrative elements and focus on practice teaching as the place where things come together.

(5) Staff Training. Develop and perform necessary staff training.

(6) Conduct the Seminar.

(7) Evaluate Against Objectives.

(8) Revise and Restructure as Necessary.

Given the once-a-year nature of Seminar and the problems of staff turnover, this will not be an easy task. Nevertheless, opportunity and necessity now demand that a good program be made better.

Examinee and Accession Quality: Past and Present

Brian K. Waters, Janice H. Laurence, and Barbara M. Means
Human Resources Research Organization

Most published research on military recruiting has focused on characteristics of accessions without considering those of examinees. To more fully understand the selection process and recruiting results, it is important to analyze not only the quality of those who enter the military but also the quality of the larger examinee group from which accessions must be drawn. This paper traces both examinee and accession test score trends since 1964 and considers the implications of the data for military manpower policy in the 1980s.

Military Enlisted Selection Process

Before examining the military test score data, it is important that the reader be familiar with the terminology used in the process of procurring recruits for the Military Services. Figure 1 graphically depicts the process.

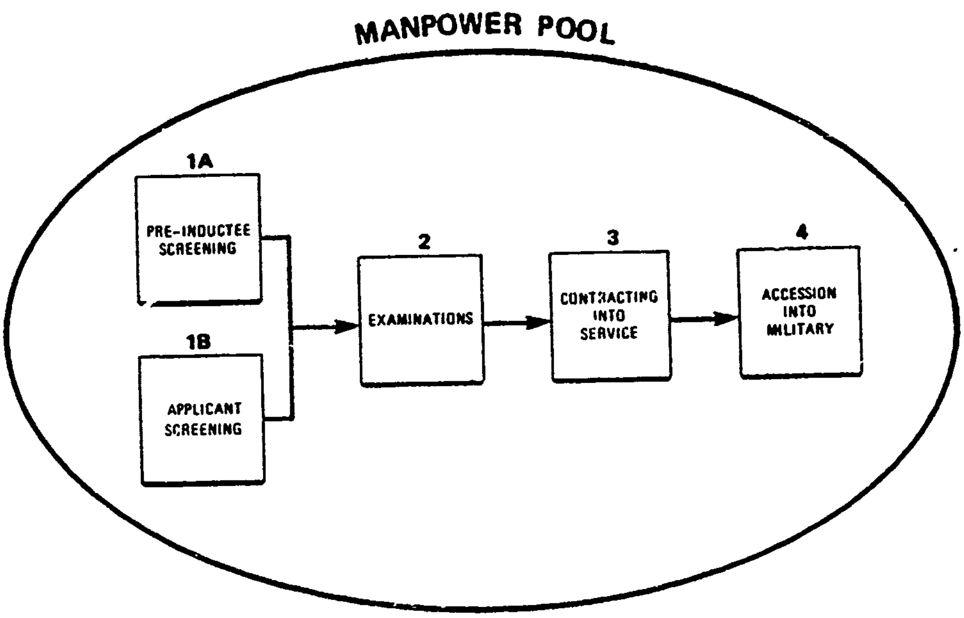


Figure 1 - **Military Enlisted Selection Process**

The manpower pool, or population from which emergency mobilization personnel would be drawn, refers to those youth, ages 17 through their late 20s, who could be called for military duty in case of a national defense emergency. Knowledge of the characteristics of this population is important to defense manpower analysts for planning purposes. From this population, as shown in step 1 of Figure 1, applicants and pre-inductees (during a draft) are screened for selection eligibility. During draft periods (the last draft call was in December 1972), local draft boards determine pre-inductee registrant eligibility. Recruiters provide the initial screening for applicants. Individuals who progress to step 2 of Figure 1, are labeled examinees. Examinees take the operational Armed Services Vocational Aptitude Battery (ASVAB) or high school testing program version of ASVAB as part of their entry procedures. Compared with the manpower pool, this examinee population is restricted both through self-selection and through initial screening by local draft boards and recruiters. (Although this paper focuses on test score data, it should be noted that additional criteria are used to determine eligibility, including level of education, physical fitness and health, citizenship, age, and moral record.) Individuals at step 3, contracts, have actually signed a contract with one of the Military Services. This category includes examinees who immediately enter the Service as well as those who

enter the delayed entry program (DEP).[1] The DEP should be kept in mind when comparing examinee data to accession data at any selected time since a relatively large subset of those who have contracted to enter the Service may not have actually entered. Not all examinees enlist. The application of selection standards and the voluntary withdrawal from the application process (or contract reneging on the part of some DEP members, reduces the number of examinees. Finally, individuals who actually enter the military (step 4) are termed accessions (or recruits). The primary goal of the recruiting process is the accession of new personnel of the quantity and the quality required to maintain authorized military strength.

Military Examinee Test Score Trends

Although there are many aspects of accession quality (including physical fitness and motivation), intellectual aptitude is the facet of quality which has received the most attention in recent years and which is the focus of this paper. Since the early 1950s intellectual aptitude or trainability has been measured by the Services using a composite of verbal and quantitative subtests from the ASVAB to compute an Armed Forces Qualification Test (AFQT) score. These scores are reported as percentiles, which have been statistically related to the performance of the mobilization population taking the aptitude test used in World War II. Hence, an individual achieving an AFQT percentile of 60 in 1982 would presumably be at the 60th percentile of the World War II mobilization group in terms of mental aptitude. Although the tests used to compute AFQT scores have changed over the years, the intent has been to hold constant the relative aptitude of an individual with a particular percentile score.

The measure of aptitude quality used here to compare examinees and accessions over the years is the proportion of each group scoring at or above the 50th percentile on the AFQT. Hence, this measure shows the proportion of accessions or examinees in a particular year who would have been in the top half of the distribution of World War II examinees in terms of intellectual aptitude. For convenience, this group will be referred to as "high quality."

Table 1 shows the proportion of male non-prior service examinees tested for entry into the Military Services between FY 1964 and FY 1981 who scored in this high-quality range. It should be noted that examinees during the draft years (1964-1973) are not entirely comparable to All-Volunteer Force (AVF) applicants since portions of the former group were draft-motivated volunteers and pre-inductee examinees. Considering the draft era and the AVF period separately, one notes a large difference in the level of examinee quality <u>between</u> the two periods and the relative consistency of examinee quality <u>within</u> each period. Although there have been large changes from year to year within each period in the number of examinees (e.g., 1,100,000 in FY 1970 vs. 650,000 in FY 1971; 466,000 in FY 1978 vs. 676,000 in FY 1981), the proportions scoring above the 50th percentile remained similar. Factors such as enlistment incentives, enlistment standards, compensation changes, and external economic trends did not have much effect upon the AFQT distributions of examinees during either the pre-AVF or the AVF period. The AVF transition, however, had an <u>enormous</u> effect upon examinee quality.

Table 1

Percent AFQT Category IIIA and Above (AFQT>50) Male Non-Prior Service Examinees by Service: 1964-1981

Fiscal Year	Percent Category I-IIIA: Army	Navy[a]	Marine Corps[a]	Air Force[a]	Total DoD
1964	39.7				41.9
1965	41.3				43.7
1966	48.0				48.2
1967	49.5				49.6
1968	47.3				47.8
1969	43.0				44.6
1970	51.4				51.0
1971	50.0				50.0
1972	49.8	50.9	33.6	55.0	49.7
1973	51.5	50.3	31.2	57.5	51.8
1974	39.6	56.3	39.3	51.6	45.1
1975	37.3	45.2	36.5	54.9	41.7
1976	32.2	39.7	40.3	42.5	36.4
1977	25.1	42.3	33.2	48.4	34.8
1978	26.5	46.5	33.7	49.8	37.4
1979	23.3	45.1	31.7	47.4	34.7
1980	23.0	50.5	36.3	50.7	37.2
1981	26.2	45.9	40.5	51.7	38.1

Sources: Data for Years 1964-1971 are based upon adjusted Preinduction Examinee Scores reported in the Office of the Surgeon General Form 1043, <u>Results of Preinduction Examinations Summary</u> and <u>Armed Forces Examining & Entrance Station Qualitative Distribution Report of Male Enlistments, Induction, and Rejections</u>, RCS DD-M(M)-663, 1964-1971. Data for years 1972-1981 were provided by the Defense Manpower Data Center.

[a]Data not available from 1964-1971.

[1]Under DEP, individuals are permitted to enlist, but not actually report for active duty for up to one year.

Given the continuation of the AVF, it would seem reasonable to assume that the proportion of high-quality examinees will remain relatively stable (in the absence of major changes in the economy or in military compensation and recruiting practices). However, the numbers of individuals in this category must be expected to decline sharply. Demographic data show that there will be a near 20% reduction in the size of the enlistment-age manpower pool during the 1980s. A drastic reduction in the number of the most desired examinees must be expected to occur within the next five to ten years. Such a forecast has considerable implications for military manpower planners.

Military Accession Test Score Trends

Although the examinee population represents the pool from which recruits must be drawn, it is the quality of the accession population which is the primary concern. Table 2 shows the percentage of male non-prior service accessions with above-average aptitude scores from FY 1972 to FY 1982.

The distributions of AFQT scores of military recruits reflect both factors under the control of, and factors independent of, military recruiting efforts and policy. Manpower requirements, incentives and compensation, recruiting and marketing resources, enlistment standards, and accession policies are examples of factors which DoD and the Congress can manipulate. But there are other important variables, such as the economy (particularly the youth unemployment rate) and attitudes toward the military, which are beyond DoD's control. Thus, accession data must be analyzed carefully. Trends rarely, if ever, reflect simple causes. Nevertheless, the AFQT scores of accessions are a prime measure of recruiting success.

Table 2

Percentage of Male Non-Prior Service Accessions Scoring in AFQT Categories I-IIIA[a] by Service from FY 1952 through FY 1982

Fiscal Year	Service: Army	Navy	Marine Corps	Air Force	Total DoD
1952	39.8	50.2	39.0	48.3	43.2
1953	42.7	56.0	42.8	54.7	45.3
1954	50.8	56.7	40.8	52.1	50.5
1955	51.9	47.0	51.8	52.0	50.6
1956	51.7	50.7	41.4	58.5	51.5
1957	48.5	51.2	46.5	59.3	52.7
1958	49.8	67.7	56.6	57.8	56.5
1959	53.0	69.2	54.5	67.1	58.8
1960	54.4	68.1	49.6	65.0	53.7
1961	56.[illegible]	69.0	58.3	62.2	50.2
1962	52.5	56.7	58.1	70.6	59.0
1963	52.1	71.8	63.5	68.2	60.5
1964	53.9	67.7	58.4	72.2	60.0
1965	52.9	65.0	60.9	69.1	59.2
1966	52.6	75.2	57.5	71.4	59.9
1967	51.5	76.1	54.2	67.5	57.5
1968	49.8	72.0	49.1	64.6	55.1
1969	51.0	56.1	46.3	64.0	55.2
1970	51.0	70.0	46.0	65.0	55.6
1971	51.0	70.8	46.7	59.3	55.0
1972	53.7	59.2	47.6	65.8	56.5
1973	53.4	59.2	48.3	58.7	57.2
1974	47.9	60.9	62.9	69.1	56.6
1975	52.8	54.6	64.6	74.4	61.1
1976	45.9	57.5	65.3	78.6	59.2
1977	32.4	55.9	47.3	72.7	47.3
1978	34.0	53.9	44.1	59.4	49.2
1979	28.5	55.4	42.6	63.6	44.1
1980	28.9	59.9	43.3	59.3	44.0
1981	39.9	61.9	54.8	67.7	54.5
1982	51.4	62.9	56.5	69.0	58.7

Sources: Data for FYs 1952-1970 were calculated from Annual Reports of the Qualitative Distribution of Military Manpower. Data for FYs 1971-1982 were provided by the Defense Manpower Data Center.

[a]Categories I-IIIA correspond to scores at or above the 50th percentile on AFQT.

One event that complicates the analysis of accession data is reflected by the numbers in Table 2. The version of the ASVAB in operation from FY 1976 through FY 1980 was miscalibrated. (The raw score calibrated as a particular percentile score was not as high as that which would have been achieved by an individual of the World War II mobilization population with the same percentile score.)

This misnorming led to the accession of individuals whose aptitude scores, after they were recalibrated, were considerably lower than assumed at the time of their original testing. (The data in Table 2 are based upon corrected scores for this period.)

Except for this five-year period, the years since the late 1950s show very consistent test score distributions. This observation suggests that the recruiting system adapts well to external changes in the recruiting market and to changes in force strength requirements. By manipulating incentives and compensation, recruiting resources, accession policy and other variables, the recruiting system has managed to hold the quality of accessions relatively constant despite marked changes in the quality of examinees.

Figure 2 displays the proportion of individuals scoring at or above the 50th percentile on the AFQT for both examinees and accessions.

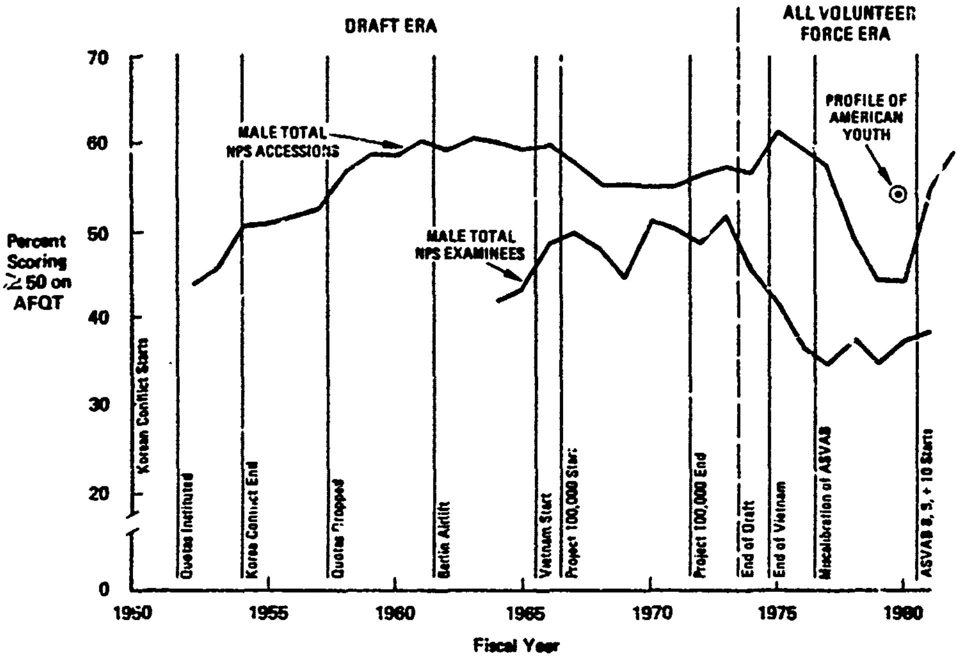


Figure 2. Percent Male Non-Prior Service Total DoD Examinees and Accessions Scoring AFQT ≥ 50: 1952-1982.

To provide an historical context for looking at these data, major events which impacted examinee and/or accession quality are indicated on the figure. Clearly the two most influential occurrences over this entire 18-year period for examinee and accession quality, respectively, were the miscalibration of the ASVAB and the advent of the AVF. The proportion of accessions scoring 50 or above on the AFQT remained between 55 and 60% from the late 1950s through the end of the draft. The increased enlistment incentives, pay, and bonuses which were provided during FY 1974 to support the transition to the AVF were clearly successful during the early AVF years. In FY 1976 the new versions of the ASVAB were introduced with miscalibrated test norms in the lower AFQT score ranges. By the time the error was verified and a new test became operational in FY 1981, the drastic drop in recruit quality shown in Figure 2 had occurred. FY 1981 and FY 1982 accession data reflect a return to more traditional quality levels.

Examinee trends[2] show two distinctly different, though relatively stable, levels of quality prior to and since the AVF transition. During the period of the Vietnam War, the proportion of high-quality examinees generally remained close to the 50% level. A significant dip in quality occurred in FYs 1968 and 1969, probably as a function of widespread avoidance of the draft and extensive college exemptions. After the redeployment of most U.S. troops from Vietnam during FY 1973 and the complete cessation of draft calls in December 1972 (mid FY 1973), examinee quality descended immediately such that approximately 40% or less scored at or above 50 on the AFQT from FY 1975 through FY 1980. FYs 1981 and 1982 show some improvement (probably attributable to high youth unemployment rates), although the AVF "base level" of high-quality examinees appears to have been established at around 36 to 40%, a considerable drop from the 50% of high-quality examinees typically experienced pre-AVF level. The authors are not aware of any other published reports documenting this clear effect of the AVF on the examinee test score distribution.

[2]The authors have been attempting to find pre-1964 examinee data reflecting AFQT category distributions for pre-inductees and/or applicants without success. Thus, Figure 2 displays data only back through 1964.

For reference, the percentage of a representative sample of young men tested in late FY 1980 for the Profile of American Youth study scoring AFQT 50 or above is included in Figure 2. As shown in Figure 2, excluding the aberrant period of the ASVAB miscalibration, DoD accessions have generally been of higher quality than the national population, although examinees have not been.

Study Implications

The gap between examinee quality and accession quality must be bridged by effective marketing, recruiting, selecting, classifying, and training of youth who comprise the male, 18-23 year old prime population for enlistment. This study suggests that, assuming continuation of the AVF through the 1980s, only about 40% of examinees will be above average in aptitude level. The number of individuals in this group is expected to decline over the next eight years by about 20% as a result of reduced birth rates. Despite this constraint, evidence from this study suggests that the AVF can work (as it did during the period from 1974 to early 1976) if sufficient resources are allocated to attract and retain quality personnel.

We would argue, as does the cartoon below, (Allison, 1982), that recent Congressional retractions of previously programmed funds seriously threaten the long-term viability of the AVF. The present recruiting market is unusually good, but pressures created by the reduced size of the manpower pool, increased technological demands of Service jobs, improved civilian youth unemployment rates, and reduced incentives to join and remain in the Services will likely result in significant losses in recruit quality during the next five years. The warning signs are loud and clear--provide the incentives for enlistment, reenlistment and career retention now or suffer the consequences in 1988.

Source: Air Force Times, 21 October 1982

Allison, R.E., Cartoon in The Air Force Times, October 21, 1982, page 21.

AD P000919

Introduction of Trade and Lifestyle Videotapes (TLVs) into a Canadian Forces Vocational Counselling Setting

Major F.P. Wilson, Lieutenant J.A. Flynn
Canadian Forces Personnel Applied Research Unit (CFPARU)

Over the past decade, the Canadian Forces (CF) have faced the serious problems of inadequate recruitment and high attrition rates. In order to learn more about these difficulties and to effect some solution, the initial counselling process for potential recruits is one area that has come under close scrutiny. Studies (Fournier & Keats, 1975; Wilson, 1980) suggested that communication between the Military Career Counsellor (MCC) and the applicant could be improved. The recruiting procedures, as they now exist, concentrate on getting information about the applicant to the MCC. However, not enough is being done to effectively provide information about the CF to the applicant. The present materials on occupational information are not sufficiently comprehensive to provide the applicant with realistic expectations about trades and lifestyle in the military. Frequently, the recruit neither has the necessary information nor the background experience required to make responsible decisions with regard to his optimum career choice.

Currently, trade and on-the-job lifestyle information is presented in printed form, using a fairly detailed technical description, augmented by brief brochures with glossy still photographs. Additionally, the MCC uses his/her own knowledge and experience to provide an account of trades and military lifestyles. Notwithstanding, a large percentage of the recruits appear to be ill prepared, as the CF lifestyles and occupations fail to match with their initial expectations. Adding to these difficulties are considerations concerning the individuals' stage of readiness to accept, understand, and internalize career information. Some career theorists (Super, 1973) describe the varying degrees of readiness on a "vocational maturity" continuum. Vocational maturity is defined as the rate and level of an individual's development with respect to career matters. Normatively, it is the congruence between an individual's vocational behaviour and the expected vocational behaviour at that age. Relative to the recruit population, who are normally in the late adolescent stage of life, the counselling tasks are both to facilitate occupational exploration and to enhance career preparation. This can only be accomplished through providing realistic, qualitative vocational information using scientifically researched communication techniques.

Applied vocational research has shown the usefulness of audiovisual presentations in aiding the process of imparting realistic and relevant occupational information. Communicating realistic occupational information by an audiovisual medium has been found to clarify job expectations by rendering them more qualitatively accurate, and has the positive long range effect of increasing job satisfaction and lowering attrition rates (Horner, Mobley, & Meglino, 1979; Ilgen, 1975; Wanous, 1975). The suggestion is that possessing accurate occupational information prior to making a decision to accept a position may lead to an increased commitment to that decision, and therefore, to a decreased probability of resignation.

Several studies have shown that visual presentations may be more effective than printed matter for those having learning difficulties, small vocabularies or difficulties in decoding written material (Gagne, 1965; Tanner, 1966). It has also been found that poor learners showed preference towards the audiovisual modality for the presentation of occupation

information (Johnson, Korn, & Dunn, 1975). In this vein, the other ranks applicants to the CF have a broad spectrum of learning abilities, education levels, and a wide range of verbal and reading comprehension skills. Hence, they should benefit from having videotaped CF trade and environmental lifestyle information available, in addition to printed information. By offering more than one information medium, recruit applicants with low levels of verbal and reading comprehension would learn more as a result of the visual presentation, while those with higher levels would be expected to learn well from any mode of information presentation. Finally, it has been demonstrated that individuals who watch vocational videotapes are more likely to be motivated to do further research and reading on career information (Fisher, 1975). Thus, in the recruiting centre context, the recruit who is exposed to occupational videotapes will more readily seek vocational information from the MCC, trade briefs, and other sources. This results in a recruit who is better informed about possible career choices, has gained in vocational maturity, and is in a better position to make important career decisions.

Development of the Trade and Lifestyle Videotapes

Five naval Trade and Lifestyle Videotapes (TLVs), depicting realistic trade information and military lifestyles, were developed by the staff of CFPARU. The TLVs take advantage of the "peer counselling" concept which holds that (in this context) an adolescent will develop a "better feel" for CF occupations and lifestyle if s/he is given that information by someone his/her own age who is already a member of the Forces and employed in that trade and environment. A series of questions was generated based on the results of occupational analyses, a survey reporting the most frequent concerns stated by prospective recruits, and a study which examined dissatisfiers of military life as indicated by current serving members (CF Occupational Analysis Report, 1980; Fournier & Keats, 1975). Tradesmen in each occupation were interviewed on film using these questions and the interviewer's voice was withdrawn during subsequent film editing. Although the tradesmen did not rehearse the answers to the questions, they were permitted to read them prior to being interviewed. It was required that each answer to a question be stated as a complete thought to obviate the necessity for the viewer to hear the question. The interviewee's remarks were then used as a voice-over describing himself or his fellow tradesmen, going about their trade tasks and general military duties, with occasional frames reverting back to the face of the interviewee. The film was subsequently transferred to videotape and edited down to a five-minute TLV.

In view of the existing evidence as to the value of audiovisual occupational presentations, and the observed inadequacy of the extant vocational counselling methods, it was decided to evaluate the communicative efficacy of videotaped trade and lifestyle information as a part of the counselling process at Canadian Forces Recruiting Centre (CFRC) Toronto. This study examines the introduction of videotapes describing the following sea trades; Weaponman Surface, Radar Plotter, Marine Engineering Mechanic, Boatswain and Signalman Sea, as well as the lifestyle at sea.

Method

Subjects

Two hundred and thirty two Anglophone recruit applicants were randomly selected from all eligible male applicants applying for other ranks trades.

Instrumentation

The TLVs described above, and brief printed summaries of the five sea trades were provided to the subjects in three of the experimental groups. The summaries were extracted from vocational material currently used by CFRCs, and thus did not precisely coincide with the information presented on the videotape. Two carrels containing an 11" colour TV monitor, video-tape player and a control panel enabled the subject to sit and watch the TLV while listening to the audio portion using a headset. A Trade and Lifestyle Inventory (TLI) was designed and pilcted to measure the amount of information learned by the subjects exposed to the two media. The inventory was made up of two parts: Section I consisted of four questions with a total of 11 parts relating to trades and lifestyle, and Section II questioned the subjects' reaction to the modality through which the trade and lifestyle information was presented.

Procedure

The subjects were randomly assigned to five groups, each containing at least 43 participants. Groups I and II acted as control groups while Group III read a printed Trade Brief (TB), Group IV watched a TLV only, and Group V watched a TLV and read a printed TB. With the exception of Group I, all Groups wrote the Pre-Test (TLI) which indicated the extent of their prior knowledge. The five Groups wrote the Post-Test (TLI), however, Groups I and II did not write Section II, the reaction portion of the inventory.

Hypotheses

1. H1. There will be no significant difference across Groups in the subjects Pre-Test scores.
2. H2. The subjects who received any treatment (Groups III, IV, V) will have significantly higher Learning scores (Post-Test minus Pre-Test) than subjects in the control Groups (Groups I, II).
3. H3 The subjects who received a bimodal treatment (Group V) will receive higher on Learning scores than (a) subjects who watched the TLVs only (Group IV) or (b) subjects who read the TBs (Group III).
4. H4. The subjects who viewed the TLVs only (Group IV) will have significantly higher Learning scores than subjects who read the TBs only (Group III).
5. H5. The subjects who scored below the Canadian Anglophone population mean on the General Classification (GC) test will show significant relatively greater improvement in Learning scores when exposed to the TLVs than those who scored above the mean.

Analysis

Quantification of Data

Prior to examining the content of the Pre- and Post-Tests, the content of the TLVs and TBs was evaluated. Using a method of non-frequency analysis (Carney, 1972), the treatments (five TLVs and five TBs) were examined by two raters who counted the number and category of concepts presented by each modality, by sea trade. This procedure provided an exhaustive list of concepts based on the contents of TLVs and TBs. Both raters then extracted the numbers of salient concepts common to both the TLVs and TBs.

Content Analysis and Reliability of Ratings

The raw information in the Pre- and Post-Tests consisted of statements written by the subjects. In order to convert these written statements into a state amenable to quantitative analysis, a non-frequency content analysis was utilized. This type of content analysis is employed to describe the content of a communication in a systematic form (Carney, 1972; Jahoda, Deutsch & Cook, 1951). Using the common concepts as criteria, two raters independently rated the Pre- and Post-Tests of 10 randomly chosen subjects who were exposed to Weaponman Surface trade information . A value of one was assigned to each concept found. Inter-Rater Reliability (IRR) was calculated for each subject rated employing the method shown below.

$$IRR = \frac{\text{no. of agreements between raters}}{\text{total number of common concepts}}$$

This process was repeated for the remaining four trades. The overall mean IRR for all trades was .85. In this manner, quantitiative values were obtained for subjects' Pre-Test, Post-Test and Learning (Post-Test minus Pre-Test) scores.

Data Analysis and Research Findings

An ANOVA model was selected to measure the main effects of the treatments. Due to the possibility of an interaction effect being caused by differences in learning ability, an analysis of covariance (ANCOVA) was computed to partial out possible variance resulting from learning ability differences as measured by the GC. The experimental model represented an Incomplete Factorial Design with unequal cell frequencies, because Groups II, III, IV and V wrote Pre- and Post-Tests, while Group I did not. The absence or presence of TLVs, TBs or Pre-Test were used as factors. An ANCOVA was carried out on Pre-Test, Post-Test, Learning scores and GC Raw scores using the General Linear Model (GLM) from the Statistical Analysis System (SAS). When the ANCOVA showed overall significance, appropriate individual a posteriori contrasts were carried out between Groups.

The covariate, GC Raw Score (because it could not be controlled for, or eliminated), was divided at the mean of the total Anglophone recruit applicant population sample, and high scoring (High GC) subjects and low scoring (Low GC) subjects were added as variables. An ANCOVA was performed on this data.

Pre-Test: An ANCOVA was performed on the four Groups of subjects' Pre-Test scores. Since the observed F value was not significant at the .05 level, there was no evidence that Pre-Test scores differed amongst the four groups. Therefore H1 was accepted and it was assumed that all subjects across Groups had approximately the same knowledge of the CF trades and lifestyle before the treatments were administered. Thus, Post-Test or Learning scores would be possible measures of knowledge increase.

Learning: An ANCOVA was performed on the four Groups of subjects' Learning scores (Post-Test minus Pre-Test score). Significance in this analyses ($F=27.06$, $p \leq .0001$) indicates that the treatments had an effect in influencing the Learning scores and H2 was accepted. Table 1

shows the mean Learning score for Groups II, III, IV, and V. It can be seen that subjects who were presented with TLVs plus TBs had a higher mean Learning score than either those presented with TLVs alone or those presented with TBs alone.

Table 1
Mean Learning Scores across Groups

Group	N	Mean Learning Scores
II	43	0.79
III	43	3.00
IV	50	5.30
V	49	5.89

A posteriori comparisons were carried out between Groups. Subjects who were presented with TLVs plus TBs (a) did not have significantly higher Learning scores than subjects presented with TLVs alone, and (b) did have significantly higher Learning scores than subjects presented with TBs alone ($F=1.34$, $p>.25$; $F=28.25$, $p\leq.0001$). Therefore, H3 (a) was rejected and H3 (b) was accepted. Learning scores for subjects who watched the TLVs only were significantly higher than subjects who read the TBs only ($F=17.86$, $p\leq.0001$). Therefore, H4 was accepted.

GC Raw Scores: GC raw scores were introduced as a covariate into the ANCOVA of Pre-Test, Post-Test and Learning scores. No significant main effect or interaction was demonstrated. This would indicate that learning ability, as measured by the GC, did not play a significant part in the study and that the treatment effectiveness was not dependent upon learning ability. Hence, H5 was rejected.

Trade and Lifestyle Inventory - Section II: Responses from Section II of thre TLI indicate subjects in Group V (TLV plus TBs) overwhelmingly preferred obtaining information by viewing the TLVs as opposed to reading the TBs. Subjects in Groups III, IV, and V were also given an opportunity to express their subjective impressions regarding the treatments. By far, more comprehensive and favourable comments were written about the TLVs in comparison to the TBs.

Results and Discussion

Strong evidence was found to suggest that presenting military trade and lifestyle information through the use of videotape causes more learning to occur than presenting essentially the same information using the traditional printed brief. Learning was measured as the difference between a Pre-Test score and a Post-Test score. Both types of treatment, TLVs and printed TBs, greatly enhanced learning over the control groups, however, Learning-scores obtained in the TLV and the TLV plus TBs groups were significantly higher than the printed TB only group. These results demonstrated that in this context, the audiovisual modality is superior to printed material in presenting vocational information.

Although it was hypothesized that participants who received the TLV plus TB treatment would score higher than both the TLV group only and the TB only, this was not found to be the case. Significantly more was learned in the TLV plus TB group than the TB only group; however, the difference in learning between the TLV plus TB group and the TLV only group was not found to be significant. Nevertheless, this finding does

not negate the possibility of the TLV and TB in concert offering a more powerful counselling tool. Material of greater complexity may require both modalities to ensure that all the information is being presented in the optimal manner. Little research has been completed examining the question of complexity of information in realistic job previews. Also, although the relationship has not been satisfactorily shown (Schnamm, 1977; Johnson, Korn & Dunn, 1975), work is ongoing to determine whether some individuals learn more from some types of media than others. These last two considerations auger strongly for maintaining both audiovisual and print at CFRCs for describing military trades and lifestyle.

GC scores were examined to determine whether the amount learned from the different treatments was contingent upon learning ability differences. Analyses yielded no evidence that learning ability, as measured by the GC, played a significant role in the amount learned from either TLVs or TBs. However, it would be expected that higher learning ability individuals would have an advantage if more complex information was being conveyed - particularly considering the close relationship between learning ability and reading comprehension. Once again, failure to establish a link between GC score and vocational information learned should not be accepted as sufficient rationale to dispense with using printed matter along with TLVs when counselling at CFRCs.

An overwhelming majority of the subjects preferred videotape to print as a trade information medium. Also, they felt that the TLVs presented a well balanced view of the CF - showing both the positive and negative aspects of military lifestyle.

References

Canadian Forces Occupational Analysis Report. Sea Operations Trades Directorate of Military Occupational Structures. April 1980.

Carney, T.F. Content Analysis: A technique for systematic inference from communication. Winnipeg. University of Manitoba Press, 1972.

Fisher, T.J. The effect of video tape model-reinforcement counselling upon the vocational information seeking behaviour of college students. Dissertation Abstract International, 1975. Aug., (2-B) 885.

Fournier, B.A., & Keates, W.E. Why do they leave? Attrition 1975 (CFPARU Working Paper 75-4). Canadian Forces Personnel Applied Research Unit, Toronto, 1975.

Gagne, R.M. The conditions of learning. New York: Holt, Rinehart and Winston, 1965.

Horner, S.O., Mobley, W.H., & Meglino, B.M. An experimental evaluation of the effects of a realistic job preview on marine recruit affect, intentions and behavior. Navy All-Volunteer Manpower R&D program of the Office of Naval Research N00014-C-0938. University of South Carolina, 1979.

Ilgen, D.R. The influence of expectations and beliefs on the motivation and adjustment of new members of military organizations. Paper presented at the "Psychology of Military Service" Conference, April 23-25, 1975, Center for Continuing Education. University of Chicago, Chicago, Illinois.

Jahoda, M., Deutsch, M., & Cook, S.W. Research methods in social relations. Part I. New York: The Dryden Press. 1951.

Johnson. W.F., Korn, T.A., & Dunn, D.J. Comparing three methods of presenting occupational information. The Vocational Guidance Quarterly, 1975, Sep., 24, (1), 62-66.

Schramm, W. Big media little media. Beverly Hills, Calif.: Sage Publications Inc., 1977.

Super, D.E. Vocational development in adolescence and early adulthood: Tasks and behaviors. In S.H. Osipow, Theories of Career Development (2nd). New York: Appleton-Century-Crofts, 1973.

Tanner, G.L. A comparative study of the efficiency of programmed instruction with seventh grade low achievers in arithmetic. Doctoral Dissertation, Univ. of Michigan, 1966.

Wanous, J.P. Tell it like it is at realistic job preview. Personnel, 1975, 52 (4), 50-60.

Wilson, F.P. Towards a more systematic counselling model for the Canadian Forces (CFPARU Working Paper 80-3). Canadian Forces Personnel Applied Research Unit, Toronto, 1980. (a).

MILITARY TESTING ASSOCIATION

24TH ANNUAL CONFERENCE

HARRY F. GREER AWARD

TO

JOE T. HAZEL

THE HARRY F. GREER AWARD IS HEREBY PRESENTED TO DR. JOE T. HAZEL OF THE AIR FORCE HUMAN RESOURCES LABORATORY. YOUR CONSISTENT SUPPORT HAS SIGNIFICANTLY CONTRIBUTED TO THE SUCCESS OF THREE MTA ANNUAL CONFERENCES.

THIS AWARD IS MADE WITH GRATITUDE AND FRIENDSHIP OF ALL ASSOCIATED WITH THE MILITARY TESTING ASSOCIATION.

INDEX OF AUTHORS

(Author: Page)

BY-LAWS OF THE MILITARY TESTING ASSOCIATION*

Article I - Name

The name of this organization shall be the Military Testing Association.

Article II - Purpose

The purpose of this Association shall be to:

A. Assemble representatives of the various armed services of the United States and such other nations as might request to discuss and exchange ideas concerning assessment of military personnel.

B. Review, study, and discuss the mission, organization, operations, and research activities of the various associated organizations engaged in military personnel assessment.

C. Foster improved personnel assessment through exploration and presentation of new techniques and procedures for behavioral measurement, occupational analysis, manpower analysis, simulation models, training programs, selection methodology, survey and feedback systems.

D. Promote cooperation in the exchange of assessment procedures, techniques and instruments.

E. Promote the assessment of military personnel as a scientific adjunct to modern military personnel management within the military and professional communities.

Article III - Participation

The following categories shall constitute membership within the MTA:

A. Primary Membership.

1. All active duty military and civilian personnel permanently assigned to an agency of the associated armed services having primary responsibility for assessment for personnel systems.

2. All civilian and active duty military personnel permanently assigned to an organization exercising direct command over an agency of the associated armed services holding primary responsibility for assessment of military personnel.

*As approved at the 1978 General Meeting of the Association, 2 Nov 78, Oklahoma City, Oklahoma

B. Associate Membership.

1. Membership in this category will be extended to permanent personnel of various governmental, educational, business, industrial and private organizations engaged in activities that parallel those of the primary membership. Associate members shall be entitled to all privileges of primary members with the exception of membership on the Steering Committee. This restriction may be waived by the majority vote of the Steering Committee.

Article IV - Dues

No annual dues shall be levied against the participants.

Article V - Steering Committee

A. The governing body of the Association shall be the Steering Committee. The Steering Committee shall consist of voting and non-voting members. Voting members are primary members of the Steering Committee. Primary membership shall include:

1. The Commanding Officers of the respective agencies of the armed services exercising responsibility for personnel assessment programs.

2. The ranking civilian professional employees of the respective agencies of the armed service exercising primary responsibility for the conduct of personnel assessment systems. Each agency shall have no more than two (2) professional civilian representatives.

B. Associate membership of the Steering Committee shall be extended by majority vote of the committee to representatives of various governmental, educational, business, industrial and private organizations whose purposes parallel those of the Association.

C. The Chairman of the Steering Committee shall be appointed by the President of the Association. The term of office shall be one year and shall begin the last day of the annual conference.

D. The Steering Committee shall have general supervision over the affairs of the Association and shall have the responsibility for all activities of the Association. The Steering Committee shall conduct the business of the Association in the interim between annual conferences of the Association by such means of communication as deemed appropriate by the President or Chairman.

E Meeting of the Steering Committee shall be held during the annual conferences of the Association and at such times as requested by the President of the Association or the Chairman of the Steering Committee. Representation from the majority of the organizations of the Steering Committee shall constitute a quorum.

Article VI - Officers

A. The officers of the Association shall consist of a President, Chairman of the Steering Committee and a Secretary.

B. The President of the Association shall be the Commanding Officer of the armed services agency coordinating the annual conference of the Association. The term of the President shall begin at the close of the annual conference of the Association and shall expire at the close of the next annual conference.

C. It shall be the duty of the President to organize and coordinate the annual conference of the Association held during his term of office, and to perform the customary duties of a President.

D. The Secretary of the Association shall be filled through appointment by the President of the Association. The term of office of the Secretary shall be the same as that of the President.

E. It shall be the duty of the Secretary of the Association to keep the records of the association, and the Steering Committee, and to conduct official correspondence of the Association, and to insure notices for conferences. The Secretary shall solicit nominations for the Harry Greer award prior to the annual conference. The Secretary shall also perform such additional duties and take such additional responsibilities as the President may delegate to him.

Article VII - Meetings

A. The Association shall hold a conference annually.

B. The annual conference of the Association shall be coordinated by the agencies of the associated armed services exercising primary responsibility for military personnel assessment. The coordinating agencies and the order of rotation will be determined annually by the Steering Committee. The coordinating agencies for at least the following three years will be announced at the annual meeting.

C. The annual conference of the Association shall be held at a time and place determined by the coordinating agency. The membership of the association shall be informed at the annual conference of the place at which the following annual conference will be held. The coordinating agency shall inform the Steering Committee of the time of the annual conference not less than six (6) months prior to the conference.

D. The coordinating agency shall exercise planning and supervision over the program of the annual conference. Final selection of program content shall be the responsibility of the coordinating organization.

E. Any other organization desiring to coordinate the conference may submit a formal request to the Chairman of the Steering Committee, no later than 18 months prior to the date they wish to serve as host.

Article VIII - Committees

A. Standing committees may be named from time to time, as required, by vote of the Steering Committee. The chairman of each standing committee shall be appointed by the Chairman of the Steering Committee. Members of standing committees shall be appointed by the Chairman of the Steering Committee in consultation with the Chairman of the committee in question. Chairmen and committee members shall serve in their appointed capacities at the discretion of the Chairman of the Steering Committee. The Chairman of the Steering Committee shall be ex officio member of all standing committees.

B. The President with the counsel and approval of the Steering Committee may appoint such ad hoc committees as are needed from time to time. An ad hoc committee shall serve until its assigned task is completed or for the length of time specified by the President in consultation with the Steering Committee.

C. All standing committees shall clear their general plans of action and new policies through the Steering Committee, and no committee or committee chairman shall enter into relationships or activities with persons or groups outside of the Association that extend beyond the approved general plan of work without the specific authorization of the Steering Committee.

D. In the interest of continuity, if any officer or member has any duty elected or appointed placed on him, and is unable to perform the designated duty, he should decline and notify at once the officers of the association that he cannot accept or continue said duty.

Article IX - Amendments

A. Amendments of these By-Laws may be made at any annual conference of the Association.

B. Amendments of the By-Laws may be made by majority vote of the assembled membership of the Association provided that the proposed amendments shall have been approved by a majority vote of the Steering Committee.

C. Proposed amendments not approved by a majority vote of the Steering Committee shall require a two-third's vote of the assembled membership of the Association.

Article X - Voting

All members in attendance shall be voting members.

Article XI - Enactment

These By-Laws shall be in force immediately upon acceptance by a majority of the assembled membership of the Association and/or amended (in force 2 November 1973).

MTA - 24TH ANNUAL CONFERENCE (1982)

SAN ANTONIO, TEXAS

STEERING COMMITTEE

Name	Rank/Title	Mailing Address
Hans Jansen	Sqn Ldr	Air Force Human Resources Laboratory AFHRL/MO Brooks AFB TX 78235
Arnold Böhrer	Commandant	Psychological Research Section CRS KAZERNE KLEIN KASTEELTJE 1000 Brussel, Belgium
Martin L. Rauch		Chief Psychologist Ministry of Defence P.O. Box 1328 53 Bonn 1 Federal Republic of Germany
R.-Eckart Rolfs	Fregatten- kapitan	Ministry of Defence - Germany Armed Forces Staff IV P.O. Box 1328 D-5300 Bonn 1 Federal Republic of Germany
Walter W. Birdsall		NAVEDTRAPRODEVCEN Code PDM Saufley Field Pensacola FL 32509
B. Michael Berger		Analysis and Evaluation Division National Headquarters Selective Service System 1023 31st Street NW Washington DC 20435
Martin Wiskoff	Dr.	Navy Personnel R&D Center San Diego CA 92152
F. J. Hawrysh		Directorate of Military Structures ATTN: DMOS National Defence Headquarters 101 Colonel By-Drive Ottawa, Ontario Canada K1A OK2

W. E. Driskill	Dr.	Air Force Occupational Measurement Center USAFOMC/OMY Randolph AFB TX 78150
J. L. Mitchell	Lt Colonel	Air Force Occupational Measurement Center USAFOMC/OMYO Randolph AFB TX 78150
D. A. Lefroy	Lt Colonel	Canadian Forces Personnel Applied Research Unit 4900 Yonge St. Toronto, Ontario, Canada M2N 6B7
John A. Burt		US Coast Guard Institute P.O. Substation 18 Oklahoma City OK 73169
Arthur C. F. Gilbert	Dr.	US Army Research Institute 5001 Eisenhower Avenue Alexandria VA 22333
Joe T. Hazel	Dr.	Air Force Human Resources Laboratory AFHRL/AZ Brooks AFB TX 78235
William C. DeBoe	Colonel	Air Force Human Resources Laboratory AFHRL/AZ Brooks AFB TX 78235

OTHER ATTENDEES

Frank L. McLanathan	Colonel (Ret)	Psychology Department St. Mary's University San Antonio TX 78284
R. O. Waldkoetter	Dr.	US Army Research Institute Field Unit P.O. Box 16117 Fort Harrison IN 46216
Shlomo Dover	Lt Col	Department for Behavioral Sciences Section Branch M.P.O.B. 01172, Israel
Stephen E. Riemer (Recorder)		Air Force Human Resources Laboratory AFHRL/AZ Brooks AFB TX 78235

NAMES AND ADDRESSES OF REGISTRANTS

WILLIAM ADAMS
CNET
N-922
NAS PENSACOLA FL 32508

HOMER W. ADKINS
US NAVY
CNET
NAS PENSACOLA FL 32508

MAJ DENIS AITKEN
ARMY SCHOOL OF TRAINING SUPPORT
RAEC CENTRE, WILTON PARK
BEACONSFIELD, BUCKS
ENGLAND

ROBERT D. ALBECK
US NAVY
CNET
NAS PENSACOLA FL 32508

WALTER G. ALBERT
AFHRL/MOM
BROOKS AFB TX 78235

MAJ ROGER W. ALFORD
AFMPC
RANDOLPH AFB TX 78150

WILLIAM E. ALLEY
AFHRL/MOD
BROOKS AFB TX 78235

DR. EARL ALLUISI
AFHRL/CCN
BROOKS AFB TX 78235

J. P. AMOR
AFHRL/MO
BROOKS AFB TX 78235

DR. MICHAEL R. ANDERSON
CGSC-CAS3
ATTN: ATZL-SWB
FORT LEAVENWORTH KS 66027

LT COL WENDELL ANDERSON
AFHRL/TS
BROOKS AFB TX 78235

DONNA C. ANGLE
US ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

THOMAS M. ANSBRO
CNET N-54
BUILDING 624
NAS PENSACOLA FL 32508

1LT CHRISTOPHER M. ANTONS
USAFOMC/OMDQA
RANDOLPH AFB TX 78150

WAYNE B. ARCHER
AFHRL/MODS
BROOKS AFB TX 78235

MAJ GAIL ARNOTT
HQ ATC/XPTTP
RANDOLPH AFB TX 78150

DR. LINDA ASLETT
USAFOMC/OMYO
RANDOLPH AFB TX 78150

DON ATKERJON
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

MAJ JOHN BALLARD
AFMPC
RANDOLPH AFB TX 78150

MAJ RODGER D. BALLENTINE
AFIT/NCSU
3400 SWIFT DRIVE
RALEIGH NC 27606

DR. CRISTINA BANKS
UNIVERSITY OF TEXAS
BEB 500, DEPT OF MANAGEMENT
AUSTIN TX 78746

LINDA BARCUS
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

DR. G. M. BARRY
FIELD RESEARCH COORDINATOR
EDUCATIONAL R&D CENTER
UNIVERSITY OF WEST FLORIDA
PENSACOLA FL 32504

THOMAS P. BECHET
TPF&C
600 THIRD AVENUE
NEW YORK NY 10016

ART BECKER
USA CHEMICAL SCHOOL
ATTN: ATZN-CM-E
FORT MCCLELLAN AL 36205

HELEN E. BELLETTI
US ARMY DEFENSE SCHOOL
FORT BLISS TX 79916

B. MICHAEL BERGER
SELECTIVE SERVICE SYSTEM
1023 31ST STREET NW
WASHINGTON DC 20435

VIRGINIA BERK
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

1LT KEVIN A BERNING
AFHRL/MOM
BROOKS AFB TX 78235

CAPT CONRAD G. BILLS
AFROTC DET. 905
6 THOMPSON HALL
WASHINGTON STATE UNIVERSITY
PULLMAN WA 99163

WALTER W. BIRDSALL
NETPDC CODE PDM
SAUFLEY FIELD
PENSACOLA FL 32509

DR. BARBARA BLACK
US ARMY RESEARCH INSTITUTE
ATTN: PERI-IK, BUILDING 2423
FORT KNOX KY 40121

DORIS E. BLACK
AFHRL/MOMA
BROOKS AFB TX 78235

DOUGLAS E. BLAIR
AFHRL/AZ
BROOKS AFB TX 78235

DR. WALLACE BLOOM
WILFORD HALL MEDICAL CENTER
LACKLAND AFB TX 78236

DR. DOUGLAS J. BOBKO
US ARMY RESEARCH INSTITUTE
ATTN: PERI-II
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

ARNOLD BOHRER
REKRUTERINGS-EN SELECTIECENRUM
SECTIE PSYCHOLOGISCH ONDERZOEK
KAZERNE KLEIN KASTEELTDJE
B 1000 BRUSSEL BELGIUM

JOHN BONDARUK
NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE MD 20755

CAPT PAUL BORDELON
AFHRL/MOMD
BROOKS AFB TX 78235

DR. ROBERT A. BOTTENBERG
AFHRL/OA
BROOKS AFB TX 78235

MAJ FREDERICK B. BOWER
AFROTC DET. 805
BOX 2620
TEXAS A&M UNIVERSITY
COLLEGE STATION TX 77843

MAJ THOMAS A. BOYNTON
AFHRL/OA
BROOKS AFB TX 78235

FRANK B. BRAUN
DATA-DESIGN LABORATORIES
1755 S JEFFERSON DAVIS HIGHWAY
ARLINGTON VA 22202

MAJ R. J. BREEN
AUSTRALIAN ARMY/FLORIDA STATE UNIV
3813 LEANE DRIVE
KILLEARN ESTATES
TALLAHASSEE FL 32308

LT DAVID BREWER
COMMANDANT (G-P-1/2)
US COAST GUARD HQ
2100 SECOND STREET SW
WASHINGTON DC 20593

DR. L. D. BROWN
TRADOC
FORT MONROE VA 23651

2LT RICHARD H. BROWN
AFHRL/MODS
BROOKS AFB TX 78235

JANICE BUCHHORN
AFHRL/TSOZ
BROOKS AFB TX 78235

DR. RUTH BUESCHER
AFHRL/AZ
BROOKS AFB TX 78235

BILLY L. BURNSIDE
US ARMY RESEARCH INSTITUTE
ATTN: STEELE HALL
FORT KNOX KY 40121

DR. BURKE BURRIGHT
AFHRL/OA
BROOKS AFB TX 78235

JOHN A. BURT
USCG INSTITUTE
PO SUBSTATION 18
OKLAHOMA CITY OK 73169

MARY-MERCEDES S. BURT
US ARMY INFANTRY SCHOOL-DOES
FORT BENNING GA 31905

LLOYD D. BURTCH
AFHRL/XR
BROOKS AFB TX 78235

DR. RONALD R. CALKINS
VOUGHT CORPORATION
PO BOX 225907
DALLAS TX 75265

ROGER L. CAMP
AFHRL/MODS (OAO CORPORATION)
BROOKS AFB TX 78235

CHARLOTTE H. CAMPBELL
HUMRRO
633 KNOX BOULEVARD
RADCLIFF KY 40160

DAVID CAMPBELL
NODAC, BUILDING 150
WASHINGTON NAVY YARD
WASHINGTON DC 20374

EDMUND J. CARBERRY
US ARMY ARMOR SCHOOL
7303 RIDAN WAY
LOUISVILLE KY 40214

ROBERT M. CARROLL
CHIEF NAVAL OPERATIONS
NAVY ARLINGTON ANNEX (OP-115C)
ROOM G-836
WASHINGTON DC 20350

CAPT TERRY D. CARROLL
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TIA
FORT SAM HOUSTON TX 78234

DR. FRANCIS A. CARTIER
DEFENSE LANGUAGE INSTITUTE
(ATFL-TD)
PRESIDIO OF MONTEREY CA 93940

DR. PATRICK CASSADY
US ARMY TRASANA
3404 JUPITER ROAD
LAS CRUCES NM 87110

DR. ALBERT CASTELLI
US MILITARY POLICE SCHOOL
ATTN: ATZN-MP-D
FORT MCCLELLAN AL 36205

ROBERT T. CHERRY
ARMY EDUCATION CENTER
ATTN: AFZA-PA-ED
FORT BRAGG NC 28307

GEORGE H. CHYNOWETH
US ARMY RETRAINING BRIGADE
FORT RILEY KS 66442

BERNARD E. CLARK
RESEARCH TRIANGLE INSTITUTE
PO BOX 12194
RESEARCH TRIANGLE PARK NC 27709

DON R. CLARK
US FORCES COMMAND
ATTN: AFIN-CBD
FORT MCPHERSON GA 30330

HARRY A. CLARK
USAFOMC
RANDOLPH AFB TX 78150

DR. HERBERT J. CLARK
AFHRL/XR
BROOKS AFB TX 78235

DR. JOHN G. CLAUDY
AMERICAN INSTITUTES FOR RESEARCH
PO BOX 1113
PALO ALTO CA 94302

JOAN M. COFIELD
USAFOMC
RANDOLPH AFB TX 78150

STEPHEN M. COLARELLI
THE BALL FOUNDATION
800 ROOSEVELT ROAD
BUILDING B, ROOM 314
GLEN ELLYN IL 60137

RICHARD A. CONNORS
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

MAJ PAUL J. COOK
433 SOUTH GROVE STREET
BOWLING GREEN OH 43402

EDNA L. COOPER
AFMTC/3250 TTW/TTS
LACKLAND AFB TX 78236

MELINDA G. COPELAND
TRW DEFENSE SYSTEMS GROUP
ONE SPACE PARK, MAIL STATION 02/1348
REDONDO BEACH CA 90278

DR. CHARLES H. CORY
NPRDC
2415 AMITY STREET
SAN DIEGO CA 92109

DOUGLAS K. COWAN
AFHRL/MODS
BROOKS AFB TX 78235

CAPT GEORGE C. CREACH
USAIS
ATTN: ATSH-I-V-ED
FORT BENNING GA 31905

DR. DALE M. DANNHAUS
USA TRASANA
785 FRANK MAES
LAS CRUCES NM 88005

LOUIS M. DATKO
AFHRL/MODS
BROOKS AFB TX 78235

DR. DOUGLASS DAVIS
CNET (N-53)
NAS PENSACOLA FL 32508

COL WILLIAM C. DEBOE
AFHRL/AZ
BROOKS AFB TX 78235

LTC HOWARD R. DECK
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

ROBERT L. DENTON
AFMPC
RANDOLPH AFB TX 78150

RICHARD W. DICKINSON
OCCUPATIONAL RESEARCH DIVISION
INDUSTRIAL ENGINEERING DEPARTMENT
TEXAS A&M UNIVERSITY
COLLEGE STATION TX 77843

DENNIS A. DILLON
INTRAN CORPORATION
4555 WEST 77TH STREET
MINNEAPOLIS MN 55435

PAUL DITULLIO
USAFOMC
RANDOLPH AFB TX 78150

MICHAEL E. DOGGETT
US NAVY
CNET
NAS PENSACOLA FL 32508

DR. JACK DOHME
ARMY RESEARCH INSTITUTE
PO BOX 476
FORT RUCKER AL 36362

1LT MARY L. DOUGHERTY
USAFATC/TTST
RANDOLPH AFB TX 78150

LTC SHLOMO DOVER
ISRAELI DEFENSE FORCES
36 RAMBAM STREET
RAANANA ISRAEL 43000

ANDREW N. DOW
CNET (N-1222)
NAS PENSACOLA FL 32508

DR. MICHAEL DRILLINGS
USARI-PERI-II
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

DR. WALTER E. DRISKILL
USAFOMC/OMY
RANDOLPH AFB TX 78150

EUGENE H. DRUCKER
HUMRRO
PO BOX 293
FORT KNOX KY 40121

CAPT PATRICIA F. DUNN
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-ZTE
FORT SAM HOUSTON TX 78234

LTC CHARLES V. DURHAM
AWC/EDV
MAXWELL AFB AL 36113

JAMES EARLES
AFHRL/MOA
BROOKS AFB TX 78235

HEINZ-JURGEN EBENRETT
GERMAN ARMED FORCES PSYCHOLOGICAL
SERVICES RESEARCH INSTITUTE
POSTFACH 205003
D-5300
BONN 2
FEDERAL REPUBLIC OF GERMANY

MAJ JOHN EDWARDS
HQ ATC/XPTT
RANDOLPH AFB TX 78150

RICHARD EDWARDS
USAFOMC
RANDOLPH AFB TX 78150

MATTHEW S. EICHINGER
AFMPC/MPCMAR
RANDOLPH AFB TX 78150

DR. RALPH EKWALL
CAS3-CGSC
ATZL-SWB
FORT LEAVENWORTH KS 66027

TIMOTHY W. ELIG
US ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

DR. JOHN A. ELLIS
NPRDC
SAN DIEGO CA 92152

FLORENCE EMERY
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

CAPT MARK S. ENGLEBRETSON
HQ ATC/XPTTP
RANDOLPH AFB TX 78150

BENJAMIN A. FAIRBANK
MCFANN-GRAY AND ASSOCIATES
5825 CALLAGHAN ROAD, SUITE 225
SAN ANTONIO TX 78228

MAJ IAN E. FALLE
USAFOMC/OMYO
RANDOLPH AFB TX 78150

DR. KENN FINSTUEN
INDIVIDUAL TRAINING DIVISION
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TII
FORT SAM HOUSTON TX 78234

1LT DAVID D. FISCHER
USAFOMC
RANDOLPH AFB TX 78150

DR. MYRON A. FISCHL
US ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

DR. EDWIN A. FLEISHMAN
ARRO
4330 EAST-WEST HIGHWAY
WASHINGTON DC 20014

ELI FLYER
HUMRRO
10 SIERRA VISTA DRIVE
MONTEREY CA 93940

PATRICK FORD
HUMRRO
633 KNOX BOULEVARD
RADCLIFF KY 40160

1LT MARTIN L. FRACKER
USAFOMC
RANDOLPH AFB TX 78150

DR. EDWARD W. FREDERICKSON
APPLIED SCIENCE ASSOCIATES
5959 GATEWAY WEST
EL PASO TX 79925

CALVIN C. FRESNE
AFHRL/TSOW
BROOKS AFB TX 78235

DR. ROBERT L. FREY
USCG (G-P-1/2/42)
WASHINGTON DC 20593

DR. BARRIE K. FRIEDMAN
RESEARCH APPLICATIONS, INC
SUITE 320 - SOUTH
1776 EAST JEFFERSON STREET
ROCKVILLE MD 20852

DAVID FRIEDMAN
RESEARCH APPLICATIONS, INC
SUITE 320 - SOUTH
1776 EAST JEFFERSON STREET
ROCKVILLE MD 20852

JAMES L. FRIEMANN
AFHRL/TSOX
BROOKS AFB TX 78235

DR. DAN FULGHAM
TECHNOLOGY, INC
300 BREESPORT
SAN ANTONIO TX 78219

DR. SIDNEY GAEL
AT&T
1776 ON THE GREEN
MORRISTOWN NJ 07960

SHARON K. GARCIA
AFHRL/MODF
BROOKS AFB TX 78235

MAJ FRANK C. GENTNER
USAFOMC
RANDOLPH AFB TX 78150

EDWARD GERAGHTY
US COAST GUARD
TRACEN GOVERNORS ISLAND
NEW YORK NY 10004

DR. J. E. GERBER
HQ VII CORPS G3 ACES DIVISION
KELLEY BARRACKS, BOX 985
APO NEW YORK 09107

GEORGE J. GERMADNIK
USAFOMC
RANDOLPH AFB TX 78150

CAPT FREDERICK W. GIBSON
USAFOMC
RANDOLPH AFB TX 78150

DR. ARTHUR C. F. GILBERT
US ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

JOYCE GIORGIA
AFHRL/MODS
BROOKS AFB TX 78235

WILLIAM GLASSCOCK
AFHRL/TS
BROOKS AFB TX 78235

M.-MERIWETHER GORDON
AFROTC/XRX
MAXWELL AFB AL 36112

MAJ CHARLES D. GORMAN
USAF/DFBL
USAF ACADEMY CO 80840

DEENE GOTT
AFHRL/MOM
BROOKS AFB TX 78235

DR. SHARRON P. GOTT
AFHRL/MODS
BROOKS AFB TX 78235

DR. BRUCE GOULD
AFHRL/MOAP
BROOKS AFB TX 78235

1LT EARL GRANT
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-ZTE
FORT SAM HOUSTON TX 78234

STEVEN D. GROFF
USAFOMC/OMT
RANDOLPH AFB TX 78150

JOSEPH H. GUERREIN
US ARMY CGSC
436 OREGON
LEAVNWORTH KS 66048

NANCY GUINN
AFHRL/MO
BROOKS AFB TX 78235

DENIS GULAKOWSKI
XMCO INC., SUITE 801
8200 GREENSBORO DRIVE
MCLEAN VA 22102

DR. WILLIAM A. HAMEL
US ARMY TRANSPORTATION SCHOOL
ATSP-ET-SFD; BUILDING 705
FORT EUSTIS VA 23606

NANCY J. HAMM
8547 RICHMOND HWY #102
ALEXANDRIA VA 22309

SAMUEL A. HARMAN
ATTN: FESA-MT-T
USAFESA
FORT BELVOIR VA 22060

JAMES H. HARRIS
HUMRRO
633 KNOX BOULEVARD
RADCLIFF KY 04160

CDR FRED J. HAWRYSH
ATTN: DMOS
NATIONAL DEFENCE HEADQUARTERS
101 COLONEL BY DRIVE
OTTAWA, ONTARIO
CANADA K2L 2A7

WILLIAM A. HAYES
CNET, CODE 01C2
NAS PENSACOLA FL 32508

DR. JOE T. HAZEL
AFHRL/AZ
BROOKS AFB TX 78235

JERRY HEDGE
AFHRL/MOA
BROOKS AFB TX 78235

SANDRAH HEIR
NCS/INTERPRETIVE SCORING SYSTEMS
PO BOX 1416
MINNEAPOLIS MN 55440

RICHARD I. HENDERSON
GEORGIA STATE UNIVERSITY
ATLANTA GA 30303

DR. ALLYN HERTZBACH
US ARMY RESEARCH INSTITUTE
ATTN: PERI-RP
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

DR. MARTHA HERZOG
DIRECTORATE OF TEST AND EVALUATION
TEST DIVISION
PRESIDIO OF MONTEREY CA 93940

EDWARD T. HEUGEL
AFHRL/AZ
BROOKS AFB TX 78235

CAPT JANICE HIGHTOWER
LEADERSHIP & MANAGEMENT
DEVELOPMENT CENTER
MAXWELL AFB AL 36112

DR. JOSEPH M. HILLERY
US POSTAL SERVICE
EMPLOYEE RELATIONS
DEPARTMENT, ROOM 9526
475 L'ENFANT PLAZA SW
WASHINGTON DC 20260

DR. EDWARD N. HOBSON
DATA-DESIGN LABORATORIES
1775 S JEFFERSON DAVIS HWY
SUITE 307
ARLINGTON VA 22702

DR. R. GENE HOFFMAN
HUMRRO
PO BOX 293
FORT KNOX KY 40121

DR. JOYCE HOGAN
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF TULSA
TULSA OK 74104

DR. ROBERT HOGAN
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF TULSA
TULSA OK 74104

JACK HOOVER
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

MAJ MICHAEL D. HOPSON
USAFOMC/OMDR
RANDOLPH AFB TX 78150

DR. CLARK L. HOSMER
39 LONGWOOD DRIVE
SHALIMAR FL 32579

BEVERLY A. HOUTZ
NAVY-HOSPITAL CORPS SCHOOL
GREAT LAKES IL 60088

JAMES B. HOWARD
US ARMY AVIATION SCHOOL
DIRECTORATE OF EVAL/STAN
FORT RUCKER AL 36362

GRACE R. HRYMOC
AFHRL/MODS
BROOKS AFB TX 78235

DR. KENNETH T. HUNT
OFFICE OF ARMOR FORCE
MANAGEMENT STANDARDIZATION
US ARMY ARMOR SCHOOL
FORT KNOX KY 40121

IAN JACKSON
PUBLIC SERVICE COMMISSION
HULL QUEBEC CANADA

WILLIAM L. JACKSON
US ARMY AVIATION CENTER
ATTN: ATZQ-TD-ES-DVCS
FORT RUCKER AL 36362

H. J. JACOB
ARMED FORCE ADMINISTER
5300 BONN 3
FEDERAL REPUBLIC OF GERMANY

JUDITH M. JACOBSON
TEMPLE UNIVERSITY
12300 ESPALIER PLACE
POTOMAC MD

SQN LDR HANS JANSEN
AFHRL/MODS
BROOKS AFB TX 78235

DR. JERRY A. JENKINS
EDUCATIONAL TESTING SERVICE
250 PIEDOMT NE, SUITE 2020
ATLANTA GA 30308

RONALD C. JENKINS
NATIONAL SECURITY AGENCY
FORT GEORGE MEADE MD 20755

HARALD E. JENSEN
INSTITUTIONAL RESEARCH HARMAN HALL
USAF ACADEMY CO 80840

ROBERT N. JOHNSON
DTD-SOLDIER SUPPORT CENTER
FORT BENJAMIN HARRISON IN 46216

WILLIAM JOHNSON
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

DR. ALAN JONES
SENIOR PSYCHOLOGIST
ROOM 426 ARCHWAY BLOCK SOUTH
SPRINGS GARDENS
LONDON SW 1
UNITED KINGDOM

LTC DOUGLAS E. JONES
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

KAREN N. JONES
US COAST GUARD INSTITUTE
PO SUBSTATION 18
OKLAHOMA CITY OK 73169

TAFT M. JOSEPH
703 SW CHAUCER CIRCLE
LAWTON OK 73505

LAURENCE T. JOSSERAND
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-ZTE
FORT SAM HOUSTON TX 78234

LT CAROLYN JUDGE
USAFOMC/OMDR
RANDOLPH AFB TX 78150

GAYLE KADURA
AFHRL/MOA
BROOKS AFB TX 78235

SRA MAUDE KAIGLER
AFHRL/MODE
BROOKS AFB TX 78235

JEFFREY E. KANTOR
AFHRL/MODE
BROOKS AFB TX 78235

JAMES B. KEETH
USAFOMC/OMYO
RANDOLPH AFB TX 78150

ROBERT S. KENNEDY
CANYON RESEARCH GROUP, INC.
PO BOX 14575
ORLANDO FL 32807

JERRY KERCE
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

CDR ROBERT H. KERR
SSO TRAINING DEVELOPMENT
CANADIAN FORCES TRAINING SYSTEM
ASTRA, ONTARIO
CANADA KOK 1B0

CAPT JAMES KIMBELL
AFHRL/XR
BROOKS AFB TX 78235

DR. BERT T. KING
OFFICE OF NAVAL RESEARCH
800 N QUINCY STREET
ARLINGTON VA 22217

DR. BENJAMIN KLEINMUNTZ
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF ILLINOIS
AT CHICAGO, BOX 4348
CHICAGO IL 60680

MAJ JAMES R. KNIGHT
USAFOMC/OMYO
RANDOLPH AFB TX 78150

LOUISE C. KNUTSON
AFHRL/TSCM
BROOKS AFB TX 78235

WILLIAM R. KOCH
MEASUREMENT & EVALUATION CENTER
UNIVERSITY OF TEXAS AT AUSTIN
BOX 7246
AUSTIN TX 78712

DR. JANOS B. KOPLYAY
AFHRL/MOM
BROOKS AFB TX 78235

DR. ARTHUR L. KOROTKIN
ARRO
4330 EAST-WEST HIGHWAY
WASHINGTON DC 20014

MARY F. KOSS
USAREUR ACES
APO NY 09701

FRED R. KULIS
US ARMY TRANSPORTATION SCHOOL
FORT EUSTIS VA 23606

DR. ANITA LANCASTER
DOD, MRA&L(MP&FM), ACCESSION POLICY
2B269
PENTAGON
WASHINGTON DC 22202

DR. DARRYL LANG
APPLIED SCIENCES ASSOCIATES
1252 COLUMBINE STREET
DENVER CO 80206

JOHN W. LANG
US NAVY
CNET
PENSACOLA FL 32508

RICHARD S. LANTERMAN
US COAST GUARD
2100 SECOND STREET SW
WASHINGTON DC 20593

JANICE LAURENCE
HUMRRO
300 N WASHINGTON STREET
ALEXANDRIA VA 22314

HARRY G. LAWRENCE
AMERICAN MANAGEMENT ENGINEERING
ASSOCIATES, INC.
1161 E COMMERCE
SAN ANTONIO TX 78205

JOHN LAWRENCE
RESEARCH TRIANLGE INSTITUTE
RESEARCH TRIANGLE PARK NC 27709

COL RAYMOND LEAHEY
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-DTD
FORT SAM HOUSTON TX 78234

LTC D. A. LEFROY
CANADIAN FORCES PERSONNEL
APPLIED CENTER
4900 YONGE STREET
TORONTO, ONTARIO
CANADA M2N 6B7

JAMES F. LEGENDRE
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

JEROME D. LEHNUS
US MILITARY ENLISTMENT
PROCESSING COMMAND
HQ MEPCOM, MEPCT-P; BLDG 83
FORT SHERIDAN IL 60037

WILLIAM LESJAK
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

SUZANNE LIPSCOMB
AFHRL/MODE
BROOKS AFB TX 78235

SCOTT W. MADOLE
AFMPC/MPCRPQZ
RANDOLPH AFB TX 78150

STEW MALCOLM
PUBLIC SERVICE COMMISSION
HULL QUEBEC CANADA

MELVIN C. MANLY
HQ ATC/TTSA
RANDOLPH AFB TX 78150

OLAN E. MARTIN
MARINE CORPS INSTITUTE
MARINE BARRACKS
EIGHTH & "EYE" STREETS SE
WASHINGTON DC 20390

RAYMOND MARTINEZ
AFHRL/TS
BROOKS AFB TX 78235

CAPT RANDY H. MASSEY
AFHRL/MOA
BROOKS AFB TX 78235

MICHAEL D. MATTHEWS
AFHRL/MODF
BROOKS AFB TX 78235

KAREN MCBEE
NETPDC-IPO2, BLDG 516
NTTC, CORRY STATION
PENSACOLA FL 32511

L. KENT MCBEE
NAVAL TECHNICAL TRAINING
CORRY STATION
PENSACOLA FL 32511

DR. BARBARA L. MCCOMBS
DENVER RESEARCH INSTITUTE
UNIVERSITY OF DENVER
DENVER CO 80208

LT TERRY MCFARLANE
USAFOMC/OMDRP
RANDOLPH AFB TX 78150

ROBERT L. MCKINLEY
AMERICAN COLLEGE TESTING
PO BOX 168
IOWA CITY IA 52243

COL F. L. MCLANATHAN
510 MARYLOUISE DRIVE
SAN ANTONIO TX 78284

KENNETH W. MCVAY
USAMMCS
ATSK-EMI
REDSTONE ARSENAL AL 35897

DOUGLAS H. MEBANE
CHIEF LEACT
DEFENSE LANGUAGE INSTITUTE
LACKLAND AFB TX 78236

DR. LARRY L. MELIZA
US ARMY RESEARCH INSTITUTE
PO BOX 5787
PRESIDIO OF MONTEREY CA 93940

DR. ISAIS T. MENCHACA
31ST TACTICAL TRAINING WING
OPERATION TRAINING DEVELOPMENT DIVISION
HOMESTEAD AFB FL 33039

CAPT HAROLD MENDES
CANADIAN FORCES APPLIED RESEARCH UNIT
4900 YONGE STREET
WILLOWDALE, ONTARIO
CANADA

ELMO E. MILLER
HUMRRO
300 N WASHINGTON STREET
ALEXANDRIA VA 22314

JEAN MITCHELL
DEFENSE LANGUAGE INSTITUTE
LACKLAND AFB TX 78236

LTC JIMMY L. MITCHELL
USAFOMC/OMYO
RANDOLPH AFB TX 78150

DR. RICHARD B. MODJESKI
ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

EDNA MORGAN
NETPDCD
BLDG 90
GREAT LAKES IL 60088

DR. JOHN E. MORRISON
US ARMY RESEARCH INSTITUTE
ATTN: PERI-IK, STEELE HALL
FORT KNOX KY 40121

JOHN D. MORROW
AIFOS/EDV
MAXWELL AFB AL 36112

CARLA JILL MULLINS
CNET, BLDG 624, CODE N-12A1
NAS PENSACOLA FL 32508

JOE MUNIZ
AFHRL/TSM
BROOKS AFB TX 78235

DANIEL J. MURAIDA
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-ZTE
FORT SAM HOUSTON TX 78234

JOHN W. MURPHY
USAIPRM
BLDG 400, ROOM 103
FORT BENJAMIN HARRISON IN 46216

MARY MYERS
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

JOE A. NARANJO
CHIEF ISD APPLICATION DIVISION
DCS/TECHNICAL TRAINING
RANDOLPH AFB TX 78150

OLIVER NELSON
HQ ATC/TTSE
RANDOLPH AFB TX 78150

DR. PETER F. NEWTON
NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE MD 20755

DR. NANCY A. NIEBOER
NAVAL POSTGRADUATE SCHOOL
504 OCEAN AVENUE
MONTEREY CA 93940

DR. GLENDA Y. NOGAMI
US ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

MARY L. NORWOOD
PERSONNEL PSYCHOLOGIST
2853 ABG WR-ALC
ROBINS AFB GA 31098

LARRY O'BRIEN
DYNAMICS RESEARCH CORPORATION
7 LOPEZ STREET
WILMINGTON MA

COL JOHN B. O'LEARY
PO 381 MAYO
UNIVERSITY OF MINNESOTA MEDICAL SCHOOL
MINNEAPOLIS MN 55455

KAREN ALEESE OLIVER
DIRECTORATE OF TRAINING DEVELOPMENT
US ARMY AIR DEFENSE SCHOOL
FORT MONROE VA 23651

MAJ MICHAEL M. OSTRANDER
USAFOMC/OMDR
RANDOLPH AFB TX 78150

DR. RONALD C. PAGE
CONTROL DATA CORPORATION
3400 WEST 66TH STREET
MINNEAPOLIS MN 55435

MIKE PARKS
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-DTD
FORT SAM HOUSTON TX 78234

SUZANNE PARRY
OSD/OUSD
5325 ESSEX CT, #151
ALEXANDRIA VA 22311

DR. JOHN J. PASS
NODAC
BLDG 150
WASHINGTON NAVY YARD
WASHINGTON DC 20374

CAPT GARY PATTERSON
USAFOMC/OMYV
RANDOLPH AFB TX 78150

VINCENT W. PATTON III
US COAST GUARD
2100 2ND STREET SW, ROOM 3505
WASHINGTON DC 20593

LTC DAVID L. PAYNE
AFHRL/MO
BROOKS AFB TX 78235

NANCY A. PERRIGO
AFHRL/TSR
BROOKS AFB TX 78235

JIM PERRY
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

GARY W. PETERSON
FLORIDA STATE UNIVERSITY
1A TULLY
TALLAHASSEE FL 32306

WILLIAM J. PHALEN
AFHRL/MOMA
BROOKS AFB TX 78235

COL ROBERT L. PHILLIPS
US ARMY
DAPE-HR
THE PENTAGON
WASHINGTON DC 20310

NETA PIKE
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

JOSEPH M. PILJAY
US ARMY ENGINEER SCHOOL
SQT-TDI-DTD-USAES
FORT BELVOIR VA 22060

MANUEL PINA
AFHRL/MOM
BROOKS AFB TX 78235

MAJ STAN POLK
AFMPC/HS
RANDOLPH AFB TX 78150

KENNETH W. POTFMPA
AFHRL/XR
BROOKS AFB TX 78235

DR. EARL H. POTTER III
DEPARTMENT OF HUMANITIES
US COAST GUARD ACADEMY
NEW LONDON CT 06320

ROBERT F. PRIEST
OFFICE OF THE DIRECTOR OF
INSTITUTIONAL RESEARCH
US MILITARY ACADEMY
WEST POINT NY 10996

DR. ROBERT D. PRITCHARD
UNIVERSITY OF HOUSTON
HOUSTON TX 77004

SUSAN E. PUCKETT
USAFOMC
RANDOLPH AFB TX 78150

LTC RONALD J. QUAYLE
INTRAN CORPORATION, SUITE 1018
4660 KENMORE AVENUE
ALEXANDRIA VA 22304

JAMES F. RAFFERTY
SENIOR EVALUATION SPECIALIST
OFFICE OF TRAINING AND DEVELOPMENT
US POSTAL SERVICE
WASHINGTON DC 20260

BELINDA D. RAMIREZ
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-TIA
FORT SAM HOUSTON TX 78234

JAMES B RASBURY
USAMMCS
ATSK-TDA
REDSTONE ARSENAL AL 35897

JOHN H. RATHKAMP
EDUC SVCS OFFICER
ARMY EDUCATION CENTER BLDG 179
FORT MCPHERSON GA 30330

MARTIN L. RAUCH
MINISTRY OF DEFENSE
POSTFACH 1328
D-5300 BONN 1
FEDERAL REPBULIC OF WEST GERMANY

MARK D. RECKASE
AMERICAN COLLEGE TESTING
PO BOX 168
IOWA CITY IA 52243

MALCOLM J. REE
AFHRL/MOAM
BROOKS AFB TX 78235

HARRIET KAY RENICK
AFHRL/OA
BROOKS AFB TX 78235

EVELYN REUELS
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-DTD
FORT SAM HOUSTON TX 78234

DIANE E. RHODES
CNET
CODE 01511
NAS PENSACOLA FL 32508

JOHN RICHARDS
DEFENSE MANPOWER DATA CENTER
300 N WASHINGTON STREET
ALEXANDRIA VA 22134

CAPT JOHN D. RICHARDS
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-ZTE
FORT SAM HOUSTON TX 78234

STEPHEN E. RIEMER
AFHRL/AZ
BROOKS AFB TX 78235

COL PAUL T. RINGENBACH
USAFOMC
RANDOLPH AFB TX 78150

BEN ROACH
AFHRL/MOAM
BROOKS AFB TX 78235

LCDR ELAINE ROBERTS
CNET
NAS PENSACOLA FL 32508

LCDR GEOFFREY P. ROBSON
ROYAL AUSTRALIAN NAVY
2963 HUNTINGTON DRIVE
TALLAHASSEE FL 32312

DR. JEROME ROHRBACH
SOUTHERN ASSOCIATION COLLEGES & SCHOOLS
795 PEACHTREE STREET NE
ATLANTA GA 30365

ECKART ROLFS
MINISTRY OF DEFENSE
ARMED FORCES STAFF
PO BOX 1328
FEDERAL REPUBLIC OF GERMANY

DR. STEPHEN S. ROOP
TEXAS TRANSPORTATION INSTITUTE
TEXAS A&M UNIVERSITY
COLLEGE STATION TX 77843

KENDALL L. ROOSE
ACADEMIC TRAINING
NAS WHITING
MILTON FL 32570

DR. PAUL ROSSMEISSL
US ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

DR. HENDRICK W. RUCK
AFHRL/MODS
BROOKS AFB TX 78235

SYDNEY SAKO
OFFICER TRAINING SCHOOL (OTS/MTCM)
LACKLAND AFB TX 78236

WILLIAM SANDS
NPRDC (CODE 12)
SAN DIEGO CA 92152

GARY G. SARLI
US ARMY RESEARCH INSTITUTE
PO BOX 6057
FORT BLISS TX 79916

DR. WORTH SCANLAND
US NAVAL EDUCATION & TRAINING
COMMAND HQS
PENSACOLA FL 32508

JOHN H. SCHAUMANN
USAFOMC
RANDOLPH AFB TX 78150

DONALD L. SCOTT
HQ ATC/TTSA
RANDOLPH AFB TX 78150

VICTOR A. SEELY
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

DR. DAVID R. SEGAL
DEPARTMENT OF SOCIOLOGY
UNIVERSITY OF MARYLAND
COLLEGE PARK MD 20742

DR. STEVE SELLMAN
OASD(MRA&L)
ROOM 2B275
THE PENTAGON
WASHINGTON DC 20301

A1C ANN SERNAK
AFHRL/MODS
BROOKS AFB TX 78235

COL PHIL SERRIN
US ARMY SOLDIER SUPPORT CENTER
FORT HARRISON IN 46216

DR. AMIEL T. SHARON
NAVSEA, SEA 072C
WASHINGTON DC 20362

DR. JAMES B. SHAW
DEPARTMENT OF PSYCHOLOGY
TEXAS A&M UNIVERSITY
COLLEGE STATION TX 77843

FAYE SHENK
USAFOMC/OMYO
RANDOLPH AFB TX 78150

LCDR WILLIAM S. SHIELDS
CANADIAN FORCES PERSONNEL
APPLIED RESEARCH UNIT
4900 YONGE STREET, #600
WILLOWDALE, ONTARIO
CANADA

MARYANN SHIPLEY
DEPARTMENT OF THE ARMY
ATTN: ATZQ-TD-ES-SAS
FORT RUCKER AL 36362

WILLIAM R. SHOEN
SENIOR ED ADVISOR
SERVICE SCHOOL COMMAND
NAVAL TRAINING CENTER
ORLANDO FL 32813

MAJ LAWRENCE O. SHORT
LMDC/ANC
MAXWELL AFB AL 36116

DR. NORMAN M. SHUMATE
US ARMY TRAINING DEVELOPMENT
FORT KNOX KY 40121

DR. WILLIE SILVA
12403 LOS CAMPOS
SAN ANTONIOS TX 78233

CAPT LEVON SIMMONS
USAFOMC
RANDOLPH AFB TX 78150

JAMES W. SINGLETARY
CNET
NAS PENSACOLA FL 32508

MARY J. SKINNER
AFHRL/MODE
BROOKS AFB TX 78235

DEWEY A. SLOUGH
NPRDC
SAN DIEGO CA 92152

BRANDON B. SMITH
UNIVERSITY OF MINNESOTA
R460 VOTECH
1954 BUFORD AVENUE
ST PAUL MN 55108

DR. ROBERT SMITH
DOES, USATSCH
ATTN: ATSP-ES
FORT EUSTIS VA 236[illegible]

MAJ ROBERT P. SMITH
HQ AFMPC/MPCRPQ
RANDOLPH AFB TX 78150

FRANK SOSA
AFHRL/OA
BROOKS AFB TX 78235

JIMMY D. SOUTER
AFHRL/TS
BROOKS AFB TX 78235

DR. DANIEL E. SPECTOR
US ARMY CHEMICAL SCHOOL

2LT ANITA SPRINGER
USAFOMC/OMYO
RANDOLPH AFB TX 78150

MICHAEL R. STALEY
AFHRL/TSOJ
BROOKS AFB TX 78235

PAUL P. STANLEY II
USAFOMC/OMD
RANDOLPH AFB TX 78150

DR. ROBERT STEEL
AFIT/LSB
WRIGHT-PATTERSON AFB OH 45433

DONALD STEPHENSON
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

DR. STAN STEPHENSON
SCHOOL OF BUSINESS
SOUTHWEST TEXAS STATE UNIVERSITY
SAN MARCOS TX 78666

DR. ROBERT STEPHENSON
AFHRL/OA
BROOKS AFB TX 78235

CDR M. M. STEVENS
NAVY OCCUPATIONAL DEVELOPMENT AND ANALYSIS CENTER, BLDG 150
WASHINGTON NAVY YARD
WASHINGTON DC 20374

JAMES R. STOKES
US COAST GUARD HQS
WASHINGTON DC 20593

A1C MICHAEL R. STOUT
AFHRL/AZ
BROOKS AFB TX 78235

JANET B. STRAFER
US ARMY ENGINE SCHOOL
SQT-TDI-DI-DTD-USAES
FORT BELVOIR VA 22060

DR. SIEGFRIED STREUFERT
DEPARTMENT OF BEHAVIORAL SCIENCE
PENNSYLVANIA STATE UNIVERSITY
HERSHEY PA 17033

MAJ WILLIAM STRICKLAND
USAFA/RRE
USAF ACADEMY CO 80840

DR. YONG SUNG
THE BALL FOUNDATION
800 ROOSEVELT ROAD
BUILDING B, ROOM 314
GLEN ELLYN IL 60137

DAVE SUTTER
USMC (CODE TAP)
WASHINGTON DC 20380

COL JOHN E. SWINDELLS
US ARMY ARMOR SCHOOL
FORT KNOX KY 40121

CAPT LISA A. TALMAGE
USMC (CODE TPI)
WASHINGTON DC 20380

MAJ RONALD W. TARR
HQ TRADOC
ATTN: ATTG-DOR
FORT MONROE VA 23651

JAY TARTELL
USAFOMC/OMYA
RANDOLPH AFB TX 78150

MSG HARVEY TAYLOR
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

STEPHEN R. TAYLOR
US ARMY INFANTRY SCHOOL
ATTN: ATSH-TP-3
FORT BENNING GA 31905

COL RONALD W. TERRY
AFHRL/CC
BROOKS AFB TX 78235

NANCY THOMPSON
OCPO/MPKEA
RANDOLPH AFB TX 78150

DELL C. TOEDT
TECHNOLOGY INC.
300 BREESPORT
PO BOX 32644
SAN ANTONIO TX 78216

JODY L. TOQUAM
PERSONNEL DECISIONS RESEARCH INSTITUTE
2415 FOSHAY TOWER
MINNEAPOLIS MN

DR. MARVIN TRATTNER
US OFFICE OF PERSONNEL MANAGEMENT
1900 E STREET NW
WASHINGTON DC 20415

MARY M. TURK
USA ORDNANCE CENTER AND SCHOOL
ABERDEEN MD 21005

DR. THOMAS C. TUTTLE
COLLEGE OF BUSINESS AND MANAGEMENT
UNIVERSITY OF MARYLAND
COLLEGE PARK MD 20742

CAPT THOMAS E. ULRICH
USAFA/DFBL
USAF ACADEMY CO 80840

DR. LONNIE D. VALENTINE
AFHRL/MOA
BROOKS AFB TX 78235

DAVID M. VANDEVOORT
ORGANIZATIONAL RESEARCH & DEVELOPMENT INC.
2455 NORTH STAR ROAD
COLUMBUS OH 43221

DAVID S. VAUGHAN
UNIVERSITY OF TEXAS AT AUSTIN
DEPARTMENT OF PSYCHOLOGY
AUSTIN TX 78712

DR. ROBERT VINEBERG
HUMRRO
27857 BERWICK DRIVE
CARMEL CA 93923

DR. LLOYD W. WADE
TEST AND EVALUATION SECTION
MARINE CORPS INSTITUTE
PO BOX 1775
WASHINGTON DC 20013

DR. HOWARD WAINER
EDUCATIONAL TESTING SERVICE
PRINCETON NJ 08541

DR. RAYMOND O. WALDKOETTER
US ARMY RESEARCH INSTITUTE
PO BOX 16117
FORT HARRISON IN 46216

WILLIAM J. WALSH
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-TIA
FORT SAM HOUSTON TX 78234

DR. JOE H. WARD
AFHRL/MPMD
BROOKS AFB TX 78235

CDR JOHN B. WASHBUSH
NAVAL ROTC UNIT
UNIVERSITY OF WISCONSIN
MADISON WI 53706

DR. BRIAN K. WATERS
HUMRRO
300 N WASHINGTON STREET
ALEXANDRIA VA 22314

THOMAS W. WATSON
AFHRL/MODF
BROOKS AFB TX 78235

DR. CHARLES N. WEAVER
AFHRL/MODF
BROOKS AFB TX 78235

ELENA J. WEBER
USAFOMC
RANDOLPH AFB TX 78150

JOSEPH L. WEEKS
AFHRL/MODS
BROOKS AFB TX 78235

JOHNNY J. WEISSMULLER
SENSIBLE SYSTEMS INC.
1702 LEE HALL
SAN ANTONIO TX 78201

MAJ JOHN R. WELSH
USAF/AFMPC
RANDOLPH AFB TX 78150

CHARLES M. WEST
CNET
NAS PENSACOLA FL 32508

CAPT LINDA WIEKHORST
AFHRL/OTA
WILLIAMS AFB AZ 85224

JAMES WILBOURN
AFHRL/MOAM
BROOKS AFB TX 78235

DR. RAND WILCOX
UCLA
145 MOORE HALL
LOS ANGELES CA 90024

JAMES L. WILDER
AWC/EDPT
MAXWELL AFB AL 36112

MAJ F. P. WILSON
CANADIAN FORCES PERSONNEL
APPLIED RESEARCH UNIT
4900 YONGE STREET, SUITE 600
WILLOWDALE, ONTARIO
CANADA

DR. LUCY B. WILSON
NATIONAL ANALYSTS
400 MARKET STREET
PHILADELPHIA PA 19106

DR. HILDA WING
US ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA VA 22333

DR. MARTIN WISKOFF
NPRDC
SAN DIEGO CA 92152

CAPT WILLIAM T. WITTMAN
USAF
LACKLAND AFB TX 78236

LARRY WOLF
AFHRL/MOD
BROOKS AFB TX 78235

DAVID V. WURST
ACADEMY OF HEALTH SCIENCES
ATTN: HSA-TDE
FORT SAM HOUSTON TX 78234

LAWRENCE ZEPEDA
ACADEMY OF HEALTH SCIENCES
ATTN: HSHA-DTD
FORT SAM HOUSTON TX 78234

HENRY L. DUBOIS
USAFOMC
RANDOLPH AFB TX 78150

JOHN HASSEN
CNET
PENSACOLA FL 32508

GEORGE R. MATHEWS III
NETPDC IPD-1
SAUFLEY FIELD
PENSACOLA FL 32509

ALBERT P. PRUETT
NETPDC IPD-1
SAUFLEY FIELD
PENSACOLA FL 32509

RAYBURN WILLIAMS
CNET
PENSACOLA FL 32508